ESSAYS ON NONPARAMETRIC ESTIMATION OF POLICY FUNCTIONS IN ASSET PRICING MODELS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Liyuan Cui

May 2017

ESSAYS ON NONPARAMETRIC ESTIMATION OF POLICY FUNCTIONS IN ASSET PRICING MODELS

Liyuan Cui, Ph.D.

Cornell University 2017

Dynamic stochastic general equilibrium (DSGE) models generally do not admit analytic solutions. Although DSGE models are widely used in macroeconomics and finance, no statistically sound estimation methods for policy functions such as the price-dividend ratio function have been developed. Because they rely on a fully specified data generating process (DGP) of state variables, numerical solution methods, extensively adopted in the literature, may discredit model evaluation due to model misspecification of state variables and poor approximations of unknown functions.

In the second chapter, I propose a convenient nonparametric 2SLS series regression method that is built on consumption based asset pricing models (CAPM), and we investigate its performance in comparison with analytic and existing numerical solutions. The new method proposes to estimate a recursively specified function embedded in Euler equations always admits a data-based closed-form solution, and it is not only easy to implement but also asymptotically free of endogeneity biases and approximation errors, even when the CAPM becomes complex. This new method does not require specifying a DGP of state variables, which avoids model misspecification of state variables and enables us to connect the solutions of DSGE models to empirical data. Our method always provides a consistent estimation of the price-dividend ratio function for a broad class of stationary Markov state variables. The

newly proposed 2SLS series regression method will become a pivotal approach for obtaining a consistent estimation of the price-dividend ratio function in the presence of a misspecified or unknown DGP of state variables, and it can help construct the most reliable and accurate model implications.

In the third chapter, I consider Dynamic stochastic general equilibrium models (DSGE) with recursive preferences, which provide powerful means for investigating the connection between economic fundamentals, asset returns and agent preferences. A system of Euler equations is derived as a pivotal tool to obtain model implications. It often involves multiple recursively specified unknown functions of state variables over different time periods, such as the price-dividend ratio function and the wealth-consumption ratio function. Given the fact that analytic solutions are extremely difficult, if not impossible, numerical solution methods for such functions are extensively adopted in the literature. Because cross dependence exists among unknown functions, all existing numerical solution methods can only provide function approximations sequentially. Therefore, approximation errors from one solution may accumulate and contaminate the others, thereby resulting in conflicting model conclusions. Despite this importance, no statistically sound methods that provide estimation and inference on this class of multiple unknown functions have been developed. Built upon the Epstein and Zin's (1989) consumption based asset pricing model (CAPM), we propose a new nonparametric generalized method of moments (GMM) series procedure and investigate its performance in comparison with existing numerical solution methods. Instead of approximating unknown functions sequentially, our method can consistently estimate all unknown functions simultaneously, while capturing their interactions using the variance-covariance of the derived estimators. Moreover, our GMM series approach is asymptotically free of simultaneous equation biases, endogeneity biases and functional form misspecification as the sample size increases, no matter how complex the DSGE model is. In addition, compared to all existing numerical solution methods which can only provide function approximations given a fully specified dynamics of state variables, our nonparametric GMM series procedure does not require any specification for the dynamics of state variables, thus avoiding potential misspecification for the data generating process (DGP) of state variables. To incorporate a wide variety of empirically relevant setups, this paper discusses two type of the GMM series estimators, namely the two-stage and continuously updating efficient (CUE) GMM series estimators. Our nonparametric CUE GMM series estimator will improve accuracy of inference when instruments are weakly correlated with Euler equation errors. Because there is an infinite number of moments due to series approximations, our nonparametric GMM series method contributes to the GMM literature by establishing a new result on consistency and asymptotic normality, which further helps facilitate rigorous inference on the DSGE model implications. Three simulation studies are considered, and our new method has been proven to perform reasonably well in the finite sample in comparison with popular numerical solution methods such as the log linearization, discretization and projection methods.

In the fourth chapter, investor extrapolation biases in the dynamics of economic fundamental variables are introduced into the traditional Lucas Jr (1978) consumption-based asset pricing models (CAPM). Given the involvement of subjec-

tive expectations in the estimation procedure, this paper proposes a feasible generalized method of moments (GMM) approach to provide consistent estimation of model parameters. Using this new estimation method, we discover different patterns of investor extrapolation biases for local investors in China, the United States, Japan and the United Kingdom. Investors in U.S., Japan and U.K. tend to react to changes in the mean levels of economic fundamentals, whereas investors in China only pay extra attention to the overall volatile levels of the aggregate economic background. Once equipped with their specific estimated extrapolation biases, models for all these four countries show good performance in explaining well-documented economic anomalies, such as the equity premium puzzle and accumulative equity returns for the aggregate stock markets. Different types of distorted investor beliefs identified in this paper help understand why China's stock market has been deviating from economic fundamentals in recent years. These distorted investor beliefs also shed light on how the regulation of China's stock market can be improved.

BIOGRAPHICAL SKETCH

Liyuan Cui was born in Beijing, China. In 2010, she received her B.S. degree in Mathematics and Finance from Wuhan University, China. After that, she obtained her M.P.S. degree in Statistics at Cornell University in 2011. She then continued her study in Economics at Cornell University and earned her Ph.D. degree in 2017.

This document is dedicated to my parents and my husband.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my chair advisor, Prof. Yongmiao Hong, for his valuable guidance and support. I greatly benefited from his vast knowledge and intelligent insight during my graduate experience. Prof. Yongmiao Hong is the one who truly made a difference in my life. It was under his influence that I developed an exuberant and earnest attitude of teaching and research.

I would like to thank my co-chair, Prof. Ming Huang, for his helpful feedback, inspiration and motivation. I must also thank the other committee members, Prof. Nimark and Prof. Tennyson for their enormous assistance, understanding and kindness through out my graduate program.

A very special thanks goes to Prof. Yongmiao Hong's wife, Xin Wang, who greatly helped me overcome numerous difficulties that I have been facing. Without her encouragement and motivation, I would not have successfully received my PhD at Cornell University.

I would like to thank my parents and my husband, Huayi Yu, for their selfish support and love through my entire life. Without their encouragement and assistance, I would not have finished this thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Solving Asset Pricing Models Via Nonparametric 2SLS Series Regression

Considerable attempts to enrich the explanatory powers of economic models have been witnessed in recent years. Economists have loaded additional factors into canonical models to enhance the understanding of well-documented economic anomalies, thereby increasing model complexity. As a result, analytic or closed-form solutions usually become extremely difficult, if not impossible, as is the case for dynamic stochastic general equilibrium (DSGE) models (Fernández-Villaverde et al., 2016). In macroeconomics and finance, Euler equations are often employed as a pivotal tool to understand the well-known equity premium puzzle. The key is to solve for price-dividend ratios as a function of state variables, which are recursively specified in Euler equations. All existing numerical solution approaches that are popularly employed to approximate the price-dividend ratio function are subject to approximation errors, which do not disappear even when the sample size goes to infinity, because they are not estimation-based procedures. In addition, for computational convenience, all currently available approaches must pre-assume an auxiliary fully specified data generating process (DGP) for state variables, which may lead to model misspecification compared to its true underlying dynamics. Therefore, one must be cautious when interpreting conclusions from DSGE models that are built on poorly approximated functions and spurious DGP of state variables. Despite their important role, solid econometric methods for estimating and inferring price-dividend ratios have not been effectively developed. This paper fills this gap in the literature by introducing a convenient nonparametric 2SLS least squares (2SLS) series regression method to estimate price-dividend ratios without requiring any knowledge of the DGP of state variables. We establish the desired asymptotic properties of the proposed method and examine its finite sample performance in comparison with popular numerical approximation methods in the literature.

The importance of accurately estimating policy functions has also been widely acknowledged in discrete choice dynamic programming problems (Bajari et al., 2007). They develop a novel method to confront dynamic equilibrium models with empirical data in both the policy function solution and structural parameter estimation stages. A reliable and flexible estimate of policy functions can enable convenient constructions of value functions and estimations of structural parameters in the second stage (Bajari et al., 2007). Dynamic programming problems, Hotz and Miller (1993) develop a method for estimating structural model parameters via matching estimated conditional choice probabilities and value functions. Bajari et al. (2007) propose nonparametrically estimating policy functions with observed optimal decision rules and state variables of imperfect competition. Among other regularity conditions, to nonparametrically estimate the policy function at each state, they need to first estimate the conditional distributions of the optimal policy under each state and construct a transition probability among different states. In their proposal of a future direction for DSGE models in macroeconomics and finance, Blanchard (2016) suggests making DSGE models less insular by fitting the data more closely and allowing the data to determine the dynamic structure of the models. However, the asset pricing literature still does not clearly understand how to connect policy function solutions with empirical data. Given observations of state variables and equilibrium conditions, this paper proposes a new method for solving asset pricing models via a convenient nonparametric 2SLS series regression procedure.

There are a number of important reasons to solve price-dividend ratios accurately. First, asset returns in each time period are functions of price-dividend ratios over different time periods. Price-dividend ratios must be solved accurately to ensure reliable conclusions about equity premiums. Second, model-implied equity returns are commonly used to obtain parameter estimates in the simulated method of moments (SMM) procedure, which was first recommended by McFadden (1989). Due to the well-known equity premium puzzle, matching sample moments of equity premiums from real empirical data with model-implied counterparts is viewed as one of the major priorities for macroeconomics and finance. Campbell and Cochrane (1999) consider the role played by consumption habit under the classical CAPM structure. Barberis et al. (1999) investigate the prospect theory on aggregate stock markets. In all these influential papers, parameter estimates are obtained by matching Euler equilibrium moments and the simulated moments of equity premiums. Therefore, potential approximation errors for the price-dividend ratio function will be incorporated into the SMM procedure, thus contaminating model parameter estimation and inference.

The distribution of model implied stock returns is directly determined by the solution of the price-dividend ratio function. Regressing these excess stock returns on the price-dividend ratio function enables forecasting analysis. Campbell (2003) demonstrates that capturing the empirical relationship between the price-dividend ratio function and excess stock returns is essential to understanding the stock market volatility puzzle. Campbell and Cochrane (1999) conduct a long-horizon regression of excess stock returns on the approximated price-dividend ratio function. They show that their model with consumption habits is able to predict the negative relationship between excess returns and stock prices. Barberis et al. (1999) study asset prices in an economy where investors are more loss averse over financial fluctuations. They solve for the price-dividend ratio function via the value function iteration method. By constructing this one-to-one mapping between state variables and price-dividend ratios, they also reveal this negative relationship between price-dividend ratios and future stock returns. Bansal and Yaron (2004) reveal a negative predictive relationship between consumption volatility and price-dividend ratios through a long-run risk model.

The price-dividend ratio function is also employed as a powerful tool for evaluating different asset pricing models. Cochrane (1992) derives the components of the variance bound of price-dividend ratios and points out that price-based volatility tests provide qualitatively different information than return-based Euler equation tests. Using the mean and variance of the price-dividend ratios as restrictions, Cliff (2001) evaluates model performance among seven popular asset pricing models. Despite improvements in current asset pricing models, it is still challenging to match the high volatility of the price-dividend ratio in the data. Using a long-run risk asset pricing model, Jagannathan and Marakani (2015) show that the variance of log price-dividend ratios can be used to identify long-run risk factors. The solution accuracy of the price-dividend ratio function also directly determines the reliability of claims about the relationship between the log price-dividend ratio and consumption growth, which has been a challenging relationship for the long-run risk asset pricing models to explain (Beeler and Campbell, 2009).

An accurate solution for the price-dividend ratio function also provides a reliable channel to test the existence of bubbles, which occurs when there is no discount rate that can explain the variance of price-dividend ratios (Cochrane, 1992). In addition, Barro (2009) conducts a different counterfactual prediction and reports that an increase in uncertainty implies a higher price-dividend ratio. The author further derives a closed-form solution for attained utility as a function of the price-dividend ratio, which enables analysis of local effects on welfare. Therefore, given all these important applications that the price-dividend ratio function assists, our paper aims to provide a convenient alternative solution method for the price-dividend ratio function which can work in a wide range of empirically relevant situations.

Our newly proposed nonparametric 2SLS series regression method can be extended without many changes to general DSGE models in macroeconomics, where estimations of model structural parameters and welfare analysis are based on the solution accuracy of policy functions. Blanchard (2016) points out that the misspecification of some part of the DSGE models will affect the estimation of model parameters. Woodford (2002) shows that the use of the log-linear approximation of policy functions such as equilibrium fluctuations in consumption, inflation and output will lead to spuriously higher expected utility under autarchy. An accurate welfare criterion via a higher-order approximation requires a characterization of policy functions with precise higher-order terms. Kim and Kim (2003) document a welfare reversal due to approximation errors of policy functions. They find that if the risk aversion is less than unity, approximation errors from the log linearization procedure will result in a spurious result regarding welfare comparisons, where autarky produces higher welfare than the complete-market economy with full risk sharing. Schmitt-Grohe and Uribe (2004) further confirm that a correct second-order approximation of the equilibrium welfare function relies on the accuracy of a second-order approximation to the policy function.

Enormous efforts have been devoted to solve DSGE models (e.g. Judd, 1992; Judd, 1998; Fernández-Villaverde and Rubio-Ramírez, 2006 and Pohl et al., 2014). Aruoba et al. (2006) and Fernández-Villaverde et al. (2016) provide a comprehensive survey of these widely used numerical solution methods. In contrast with the current numerical solution literature, our newly proposed method does not require modelling and estimating the conditional distributions of state variables over time, thereby avoiding possible model implications due to misspecified DGP of state variables. The importance of capturing the true properties of time series variables in obtaining reliable model implications has been emphasized by Caballero (1990), Browning et al. (1999) and Meghir and Pistaferri (2004). Due to its computational convenience, the independent and identical distribution (IID) assumption on consumption and income innovations has been widely used in the literature. However, the IID assumption is too restrictive to reflect the true stochastic processes and can lead to potentially wrong model implications. Caballero (1990) point out the potential benefits and the importance of relaxing IID assumptions when explaining the excess smoothness and the excess sensitivity of consumption to unanticipated and anticipated labor-income changes. Allowing for additional unobserved transitory and permanent effects in the consumption growth process, Banks et al. (2001) use an ARCH framework for modelling the time varying income risk process instead of the IID assumption. They find that time variation in the risk components is the key factor that drives the precautionary savings effect. Meghir and Pistaferri (2004) find that model misspecification on the stochastic process of income innovations can lead to wrong conclusions about the effect of individual behavior on consumption decisions. Fernández-Villaverde and Rubio-Ramírez (2007) find strong evidence of stochastic variances in the U.S. aggregate time series. In this paper, we further contribute to the literature by removing not only the IID assumption but also all distributional assumptions on the innovation processes of state variables.

Our paper also relaxes the widely used linear time series assumption made on the DGP of state variables. Because current numerical solution methods require a fully specified conditional density for computational convenience, the AR(1) process is commonly use in both macroeconomics and finance. However, Cecchetti et al. (2000) find evidence that a threshold model that reflects different evolutions of U.S. aggregate consumption growth can help explain the equity premium puzzle. Our paper works for both linear and nonlinear time series, and can capture the role played by nonlinear features of state variables in policy functions.

Unlike methods in the current literature, the newly proposed functional estimation method in our paper does not rely on any assumption on conditional densities. It significantly facilitates policy function and model structural parameter estimations. Because most DSGE models do not offer analytic or numerically available likelihood functions, much progress has been made to reduce errors in approximated likelihood functions. Fernández-Villaverde and Rubio-Ramírez (2007) document the difference between the approximated likelihood function derived from linearized Euler equations and exact ones and introduce a helpful tool to evaluate the likelihood functions in both linear and nonlinear macroeconomic models via particle filtering. To ensure the validity and asymptotic efficiency of particle filtering, distributional assumptions are required for both the innovation processes of state variables and measurement errors. Gallant and Tauchen (1996) avoid approximating likelihood functions by choosing from some auxiliary models. However, it is difficult in practice to find the most appropriate set of auxiliary models. In addition, both procedures are computationally intensive. To circumvent these difficulties, our paper proposes an alternative functional estimation method that yields consistent and closed-form solutions for policy functions without involving conditional densities and auxiliary candidate models.

Chen et al. (2013) also propose a way to incorporates empirical data into the first stage before estimating model parameters. It is accomplished through a nonparametric estimation of the ratio between the value function and consumption, which is also a function of state variables. However, their method does not provide a direct link to either model implied stock returns or price-dividend ratios, and therefore does not assist in forecasting analysis. In addition, it does not aid the creation of a time series of simulated data, which could be used to construct comparisons between model implied asset returns and empirical counterparts. Our paper provides an alternative solution method that easily satisfies all these practical requirements.

Throughout this paper, we consider a class of time-separable CAPMs with rational expectations in an endowment economy. This class of models is recognized as a cornerstone of asset pricing theory. Campbell and Cochrane (2000) point out that most asset pricing models can be derived as various specifications under this CAPM framework. Therefore, we use this general class of CAPMs as the basis for our nonparametric 2SLS series regression method. We represent the Euler equation equivalently as a nonlinear time series regression model, where the regression function contains the unknown price-dividend ratio function over two time periods. This kind of recursive specification suffers from endogeneity, which would lead to estimation biases if a nonparametric ordinary least squares (OLS) series method were used. To tackle the endogeneity problem, we propose a convenient nonparametric 2SLS regression method to estimate the price-dividend ratio function.

The projection method can also be interpreted as a nonlinear OLS procedure, which provides a global solution for the price-dividend ratio function if the weighting function is equal to the first order derivative of the Euler equation errors with respect to projection coefficients (Fernández-Villaverde et al., 2016). However, this method is essentially different from our newly proposed 2SLS series regression method. The OLS implementation of the projection method requires a prefixed order for the series expansions and involves the calculation of conditional expectations with a fully specified conditional density function. Therefore, the system of equations is usually difficult to solve, and does not guarantee a closed-form solution. Our new method is asymptotically free of endogeneity biases and functional form misspecification, and has a convenient data-based closed-form solution no matter how complex the DSGE model is. It can perform estimation and evaluation for the class of CAPMs without having to specify the data generating process (DGP) for stationary Markov state variables. We note that in a different context, Newey and Powell (2003) first proposed the idea of nonparametric 2SLS series estimation using ridge regressions.

Following Burnside (1998), Tsionas (2003) and Calin et al. (2005), we compare the accuracy of the nonparametric 2SLS series regression method with analytic solutions under some special circumstances. Analytic solutions, when available, are the best benchmark to evaluate finite sample performance. Distances between numerical solution approaches and analytic results can be captured visually. We examine two CAPM setups, namely Mehra and Prescott's (1985) model and Campbell and Cochrane's (1999) consumption habit model in situations where the dynamics of state variables are fully correctly specified and misspecified respectively. The first CAPM has traceable analytic solutions under specific circumstances (Burnside, 1998). The second CAPM is prevalent because of its theoretical contribution. Since this model does not have analytic solutions, we adopt it as an example to discuss model evaluation based on different approximation solution methods. We find that our nonparametric 2SLS series regression method performs substantially well in both scenarios given its reasonable and robust performance over various parametrizations and its broad application.

1.2 A Nonparametric GMM Series Approach to Solving Multi-equation Asset Pricing Models with Recursive Preferences

Considerable attempts to enrich the explanatory powers of economic models have been witnessed in recent years. Economists load additional factors into canonical models to enhance the understanding of well-documented economic anomalies, thereby increasing model complexity. In doing so, analytical or closed-form solutions usually become extremely difficult, if not impossible, which is especially true for dynamic general stochastic equilibrium (DSGE) models. Numerical methods are widely applied to obtain model implications. In macroeconomics and finance, Euler equations are derived as a pivotal tool to understand the well-known equity premium puzzles. Price-dividend ratios are specified recursively in Euler equations under rational expectations. The solution of price-dividend ratios determines asset returns and model evaluations. Despite this important role, econometric methods for estimating and inferring price-dividend ratios are not effectively developed. In addition, current popularly adopted numerical approaches all suffer from misspecification or approximation errors, which do not disappear even when the sample size goes to infinity. Cautions must be practiced when interpreting results built on illy-specified functions. This paper fills this gap in the literature by introducing an instrumental variable (IV) nonparametric two-stage series regression method to estimate price-dividend ratios. We establish the asymptotic properties of the proposed method and examine its finite sample performance in comparison with popular analytical approximation methods in the literature.

There are a number of important reasons to solve price-dividend ratios accurately. First, asset returns in each time period t are functions of price-dividend ratios in times t and $t-1$. Price-dividend ratios must be solved accurately to ensure reliable conclusions on equity premiums. Second, model-implied equity returns are commonly used to obtain parameter estimates in the simulated method of moments (SMM) procedure. Due to the so-called equity premium puzzle, matching the sample moments of the equity premiums from real empirical data with model-implied counter-parters is viewed as one of the major priorities in the macroeconomics and finance literature. In this line of the literature, Campbell and Cochrane (1999) consider the role played by consumption habit under the traditional CAPM structure. Barberis et al. (1999) investigate the prospect theory on aggregate stock markets. Cecchetti et al. (2000) address the inconsistency between the subjective and objective expectations in the CAPM framework. In all these influential papers, parameter estimates are obtained by including both Euler equilibrium moments and simulated moments of equity premiums. Approximation or misspecification errors from price-dividend ratios are then incorporated in the SMM procedure, thereby contaminating estimation results and model inferences.

We consider a class of time-separable CAPM with rational expectations in an endowment economy throughout this paper. This class of models is recognized as a cornerstone in the asset pricing theory. Campbell and Cochrane (2000) point out that most asset pricing models are derived as specifications under this CAPM framework. Therefore, we use the class of CAPM's as the basis to introduce our IV nonparametric two-stage series regression method. We treat the Euler equation equivalently as a time series nonlinear regression model by adding unobservable aggregate pricing shocks. The nonlinear regression functions are composed of unknown functions of interest (i.e., price-dividend ratios) over two time periods. This kind of recursive specification generates endogeneity, which would lead to estimation biases if a nonparametric least squares estimation method were used. In a different content, Newey and Powell (2003) first propose the idea of nonparametric two-stage least-squares series estimation method using ridge regressions. Inspired by Hong and White (1995) and Newey and Powell (2003), we employ an instrumental variable nonparametric two-stage series regression method to estimate price-dividend ratios. This new method is asymptotically free of misspecification errors. It can perform estimation and evaluation for the class of CAPM's in a wide variety of empirically relevant setups.

Following Burnside (1998), Tsionas (2003) and Calin et al. (2005), we can compare the accuracy of our IV two-stage series regression method with popular numerical methods and analytic solutions in the literature. Analytical solutions are the best benchmark to evaluate finite sample performance. Distances between numerical approaches and analytical results can be captured visually. We examine two CAPM setups, namely Mehra and Prescott's (1985) standard CAPM and Campbell and Cochrane's (1995) consumption habit CAPM. The first CAPM has traceable analytic solutions under specific circumstances (Burnside, 1998). The second CAPM is prevalent because of its theoretical contributions. Since this model does not have analytic solutions. we adopt it as an example to discuss model evaluations based on different approximation solution methods. The computational convenience of the IV nonparametric series estimation method is not affected by the model complexity. Our empirically relevant simulation studies show that our IV two-stage series regression method outperforms the existing popular solution methods such as permutation, projection and the VFI method, especially for highly nonlinear functions of price-dividend ratios and for tail areas. We conclude that our IV two-stage series regression method performs the best in estimating the price-dividend ratios because of its stable and robust performance over various parametrization.

1.3 Extrapolation Bias in Economic Fundamentals and the Aggregate Stock Market

The Chinese government aims to build an interactive bilateral relationship between its stock market and the real economy. From the government's point, the stock market should play a crucial role in serving the real economy. Voluminous studies have explored the stock market of the United States, the United Kingdom and Japan, and found the existence of a strong equilibrium between the stock market and the real economy (Fama, 1990; Cheung and Ng, 1998). In comparison, this fact is different in China. Since 1990, China has been experiencing rapid development in many aspects. Unfortunately, China's stock market does not seem as glorious as its real economy. The performance of China's stock market has been contradicting with investor performance and government's expectations, while deviating from the real economy. Han and Hong (2014) study a famous Chinese industry policy implemented in 2010 and find that China's stock market failed to deliver enough support to the real economy. Specifically, 4 trillion RMB was injected into China's real economy, but only a few related industries received visible feedback from the stock market. The Chinese government is currently facing a thorny and essential task. First, why does China's stock market fail to react positively to the real economy? Second, how should China's stock market be reformed to ensure that it can better facilitate the real economy?

Taylor and Tonks (1989) point out the view that major stock markets of the world are converging at least over the long-run. Fraser and Oyefeso (2005) suggest that a single common stochastic growth component binds national equity market together. Except for fundamental variables that exhibit influence on the stock markets, investor behaviors are another important factor that affects equity markets. A remarkable similarity between Japanese and U.S. institutional investors in a number of attitudinal and behavioral dimensions is reported by Shiller et al. (1991). One of the most important media between the stock market and real economy is investors.

Investors come from different walks of life, assume various positions and span different ages. Investors are affected differently by the real economy, public news, and policies. Given these characteristics, they tend to form their own beliefs about the real economy. The question raised and answered in our paper is whether or not investors in China, the United States, Japan and the United Kingdom hold homogeneous beliefs in fundamental variables, and whether or not they act on their beliefs when trading in the stock market. To our best knowledge, this paper is among the first that quantifies the differences of investor beliefs cross countries.

A well-known survey conducted by Greenwood and Shleifer (2014) reports that investors act on their distorted beliefs, even in the United States where the market is mature. Investors' expectations of stock returns differ dramatically from the expected returns predicted from consumption-based asset pricing models. Investors tend to extrapolate historical stock prices when forecasting future stock performance. Traditional assumptions on investor behaviours assume that investors are rational in all aspects and that they can correctly perceive all relevant operating mechanisms in the stock market. This survey finding casts serious doubt on the fully rational assumptions imposed in CAPM models. Economists are concerned with the possibility that participants in the survey of Greenwood and Shleifer (2014) may answer questions based on their understanding of economic fundamentals instead of stock prices. Cui (2016) study the stock market of the United States from 18901990, which enabled the author to explore this possibility by allowing extrapolation biases on economic fundamentals; such extrapolation helps explain several economic anomalies, such as the equity premium puzzle. Our paper further explores this question by studying the modern stock market in China from 20022015.

China's stock market, established in 1996, is far younger and undeveloped compared with that of the United States. Individual investors are the major component of China's stock market, comprising almost 90% of the total stock market participants in China. Meanwhile, institutional and professional traders comprise the United States' stock market. Given this structure, arbitrage opportunities are not rare in China's stock markets. During trading, China's individual investors tend to follow the majority without applying scientific calculations. These facts make China's stock market irrational (Han and Hong, 2014). Furthermore, investors from these two countries receive utilities in different ways. In the present paper, we find that investors from these two countries have extrapolation biases on economic fundamentals, but such investors behave in different ways. Investors from the United States tend to be sensitive to changes in the values of economic fundamentals. By contrast, China's investors are concerned about the volatility of the economic fundamentals used in their stock trading strategies.

The subjective beliefs of investors deviate from the objective ones because investors are assumed to be rational in all aspects, except when holding extrapolation biases on economic fundamentals. This situation leads to the appearance of subjective expectations in Euler equations. Such subjective expectations make the generalized method of moments (GMM) estimation inappropriate because it is built on mathematical expectations. In the current study, we propose a method for adopting GMM in a framework with subjective expectations. This approach enables future relevant studies to conduct estimations and statistical inference by incorporating additional psychological evidence into the asset pricing literature.

CHAPTER 2

SOLVING ASSET PRICING MODELS VIA NONPARAMETRIC 2SLS SERIES REGRESSION

2.1 Framework

An investor's optimal decision rule in optimizing the expected life-time utility gives the following basic consumption-based model as (Cochrane, 2009):

$$f_t = E[m(X_{t+1})(f_{t+1} + 1)|I_t], \tag{2.1}$$

where X_t is the state variable that summarizes the law of motions in the system, $m(X_t)$ is a known function of the model-specific stochastic discount factor (SDF) and state variable X_t, I_t denotes all information available at time t, P_t is the price of the risky asset at time t, D_t is the dividend payments, and $f_t \equiv \frac{P_t}{D_t}$ is the unknown price-dividend ratio function. Without loss of generality, we assume that $E(\cdot|I_t)$ is the rational expectation, which coincides with the mathematical conditional expectation [1]. Hansen and Scheinkman (2012) point out the generality of Markov processes in CAPMs. Therefore, our basic objective is to provide a data-based and closed-form consistent estimate of the price-dividend ratio function f_t embedded in Euler equation (2.1) at each time period t for a wide class of Markov and stationary state variables X_t.

[1] Non-model consistent expectations occur when subjective expectations differ from objective expectations. We can convert the subjective expectation back to the mathematical one by the Radyon-Nikodym theory

In Mehra and Prescott's (1985) framework, there is one infinitely-lived representative agent in an endowment economy. There are one risky asset with 1 net supply and one risk-free asset with 0 net supply in equilibrium. The agent will maximize the expected life-time utility at time zero, namely

$$\begin{aligned} &\max_{\{C_t\}} E \sum_{t=0}^{\infty} \beta^{t-1} \frac{C_t^{1-\gamma}}{1-\gamma} \\ &\text{s.t.}\, C_t + P_{t+1}\theta_{t+1} + Q_t b_{t+1} = b_t + (D_t + P_t)\theta_t, \end{aligned} \tag{2.2}$$

where $X_t = ln(C_t/C_{t-1})$, $X_{t+1} - \mu = \Gamma(X_t - \mu) + u_{t+1}$, and $u_{t+1} \sim IIDN(0, \sigma_u^2)$. C_t is the consumption at time t, β is the time discount factor, Q_t is the price of the risk-free asset at time t which pays off 1 at time $t+1$, b_{t+1} is the position of the risk-free asset, and θ_{t+1} is the share of the risky asset at time $t+1$. The known function $m(X_{t+1}) \equiv \beta e^{(1-\gamma)X_{t+1}}$. Given the dynamics of the driving vector $\{X_t\}$ and the model structural parameters, the solution of price-dividend ratios $\{f_t\}$, which is specified recursively in Equation (2.1), is of our central interest.

Analytic solutions for price-dividend ratios $\{f_t\}$ are rather difficult even for this simplest CAPM. Economists have been seeking solutions for asset pricing models since the seminal work by Mehra and Prescott (1985). Because a moderate difference in asset prices can change an investor's decision regarding utility maximization, Burnside (1998) derives sufficient conditions for the existence of analytic solutions of price-dividend ratios. It requires time-separable preference functions and Gaussian shocks with certain restrictive inequality conditions. Burnside (1998) proves that only under the condition $r \equiv \beta e^{(1-\gamma)\mu + \frac{1}{2}(1-\gamma)^2 \frac{\sigma^2}{(1-\Gamma)^2}} < 1$ is there an analytic solution for price-dividend ratios $\{f_t\}$. Tsionas (2003) relaxes the above restrictions to a stationary bounded non-Gaussian process with time-separable utilities. Despite this improvement, analytic solutions for price-dividend ratios $\{f_t\}$ are not guaranteed in a full set of the environment. When the CAPM becomes complex, more restrictive conditions must be imposed. For example, Calin et al. (2005) provide sufficient conditions to analytically solve Abel's (1990) model. Price-dividend ratios are assumed to be a convergent power series near every point of an open interval and have a holomorphic pricing kernel. The state variable is strictly restricted to a one-dimensional Markov process. Unfortunately, it is impossible to perform model evaluation over the entire parameter space even using the analytic solution of Calin et al. (2005), because Calin et al.'s (2005) conditions for the existence of a unique solution to Campbell and Cochrane's (1999) model in the real vector space of all continuous functions do not hold.

Most DSGE models facing difficulties in analytic solutions turn to numerical approximation approaches. Widely used numerical solutions include the discretization, perturbation and projection methods. The discretization method is accomplished by exactly solving a finite number of points within a support and interpolating the areas between grid points. This method was first applied in solving the social planner's problem of a stochastic neoclassical growth model by Christiano (1990). Although various interpolation methods (e.g., linear and cubic interpolations) have been introduced, the discretization method still suffers from interpolation biases, which do not disappear when the sample size goes to infinity. Furthermore, given the algorithm of the discretization method, the rate of convergence and approximation goodness depend on the value of model parameters and the curvature of the unknown function. Chen et al. (2008) point out that several model implications from this method adopted by Campbell and Cochrane (1999) fail to hold when price-dividend ratios $\{f_t\}$ are solved analytically using the complex theory with some special parameter values. In contrast, the novel estimation methodology proposed in this paper enables consistent estimation of the price-dividend ratio function for the entire support and whole distribution of state variables, which also avoids interpolation biases when the sample size grows.

The perturbation method is popular because of its wide applications and computational convenience. The essence of this method is Taylor's theory. Dating back to the 19th century, the perturbation method has been popularly used in physics and other natural sciences. It was first popularized in economics by Judd (1998). A pre-specified functional form is obtained by expanding the price-dividend ratio function around certain steady states. However, there is stilll heated debate around the judgement of steady points (Juillard, 2011). Furthermore, the perturbation method is challenged by approximation errors, regardless of the choices of steady states. Because it needs extra effort to compute partial derivatives of Euler equations up to a higher order p, a popular approach is to linearize the model around some steady states. Using a canonical stochastic growth model, Aruoba et al. (2006) find that the performance of the linearization method is disappointing in many respects. While the linearization method is computationally fast and can obtain reasonable solutions for simple functions, the approximation errors become substantially larger for complex models, and do not vanish as the sample size increases because they are not estimation-based. What's more, when the state variables are discrete or Euler equations are non-differentiable, the perturbation method will not be of any use. The newly proposed instrumental variable nonparametric 2SLS series regression method works with both continuous and discrete state variables, and does not involve computations of partial derivatives. As the sample size goes, the newly proposed method is asymptotically free of approximation errors no matter how complex the model is.

First introduced by Judd (1992), the projection method is appealing due to its global approximation in the entire domain. It delivers an approximation without involving additional interpolation techniques. The issue is that an appropriate polynomial order p must be specified as a priori. Furthermore, the boundary regions of state variables may become too wide when the dynamics of state variables have high persistence in absolute value. This will result in a loss of accuracy in the projection method (Culham, 2005). Furthermore, Santos (2000) shows how changes in the curvature of the utility function and the time discount rate can influence the size of Euler equation errors and therefore bound the approximation errors of numerical solution methods.

These numerical solution methods are all extensively used in the literature because of their wide scope of application, weaker restriction and ease of computation. However, deciding which one performs the best is difficult because pros and cons accompany all of these methods (Culham, 2005). Taylor and Uhlig (1990) show that even for the simple growth model, different numerical solution techniques may display various results for the model. Den Haan and Marcet (1994) reach an important conclusion that numerical solution methods cannot be used interchangeably in general. In addition, one of the most commonly used measures for goodness of approximation is the relative error, which is defined as the approximated Euler equation divided by the approximated price-dividend ratio function. However, Calin et al. (2005) point out that relative errors do not necessarily reflect the accuracy of price-dividend ratios.

What is more, due to computational concerns, all current solution methods for the price-dividend ratio function described in Euler equations (2.1) require some auxiliary and artificial assumptions on the conditional distribution of X_{t+1} given X_t and its innovation processes. For example, Mehra and Prescott (1985) consider an AR(1) distributed state variable. Cecchetti et al. (2000) specify a threshold model for the state variable. The driving factor (i.e., the state variable) in Campbell and Cochrane (1999) is assumed to follow an AR(1) process with heteroskedasticity. Despite the substantial progress that has been made in the development of more realistic and reasonable DGPs, all of the auxiliary DGPs made on state variables may not capture their true underlying processes, thereby severely distorting model implications. To take the DSGE models directly to real data in a rigorous and convenient way, this paper proposes a data-based 2SLS series regression procedure, which does not require any prior knowledge of the DGP of state variables, therefore avoiding approximation errors due to a misspecified DGP of state variables.

From an econometric perspective, all existing popular numerical approximation approaches are equivalent to various parametric models for price-dividend ratios, where an approximating functional form is pre-specified as a prior. There is no assurance that a parametric model which is chosen for analytic or computational convenience will contain the true price-dividend ratio function or even a good approximation of it. Therefore, these parametric approximations can cause misleading inferences about and judgements of model performance due to potential approximation errors. It is important to provide an uniformly accurate numerical solution for the price-dividend ratio function f_t under various empirically relevant setups. A large proportion of macroeconomics and financial models are built upon stationary Markov processes for state variables. In the DSGE literature, it has been a convention to work with stationary Markov state variables since Hansen and Singleton (1982), Gallant and Tauchen (1989) and Hansen and Scheinkman (2012). We shall allow serial dependence under Markov processes. Our proposed method of the price-dividend ratio function is built on stationary, Markov, non-Gaussian and multivariate situations without imposing any parametric specification on the DGP of state vector X_t. The most important feature of our nonparametric 2SLS series regression approach is that we use a nonparametric model for the price-dividend ratio function f_t that can be consistently estimated from observed data and so is free of functional form misspecification when the sample size goes to infinity. At the same time, this new method yields a convenient closed-form solution, no matter how complex the DSGE model is.

To consistently estimate the recursively specified unknown function f_t in the Euler equation described in Equation (2.1), it is essential to establish the existence and uniqueness of the solution f_t^o. Using the linearity property of expectations, Equation

(2.1) can be equivalently expressed as an integral equation of the second kind:

$$f(X_t) - \int K(X_t, X_{t+1}) f(X_{t+1}) dX_{t+1} = f(X_t) - (Af)(X_t) = \pi_t, \qquad (2.3)$$

where $\pi_t = E[m(X_{t+1})f(X_{t+1})|X_t]$, $K(X_t, X_{t+1}) = m(X_{t+1})g(X_{t+1}|X_t)$ is the kernel of an integral operator A, and $g(X_{t+1}|X_t)$ is the conditional density of X_{t+1} given X_t. The linear operator A: $X \to X$ on a normed space $\mathbf{X}$ with $G \subset X$ is defined by:

$$(Af)(X_t) \equiv \int_G K(X_t, X_{t+1}) f(X_{t+1}) dX_{t+1}, \qquad (2.4)$$

The following conditions ensure the existence of a unique solution in Equation (2.1) or (2.3).

Assumption 2.1.1. *$G \in \mathbf{X}$ has finite-dimensional range $A(G)$.*

Assumption 2.1.2. *The linear integral operator A is bounded.*

Assumption 2.1.3. *The homogenous equation $f - Af = 0$ only has the trivial solution $f = 0$.*

Assumption 3.2.3 requires a finite number of state variables. A linear operator is bounded if there exists a positive number C so that $||Af|| \leq C||f||$. Assumptions 3.2.3 and 2.1.2 imply that linear operator A is compact on a normed space $\mathbf{X}$. If our interested domain G is Jordan measurable so that its characteristic function is Riemann integrable, the linear operator A will be bounded with a continuous kernel.

Theorem 2.1.1. *Under Assumptions 3.2.3-3.3.1, Equation (2.3) has a unique solution $f^o \in \mathbf{X}$.*

After establishing the existence and uniqueness of the solution for the recursively specified price-dividend ratios, we shall estimate this unique solution $f_t^o \equiv f^o(X_t)$ using an nonparametric 2SLS series regression method.

2.2 Nonparametric 2SLS Series Regression

We can represent the Euler equation equivalently as a nonlinear time series regression model. Let $F_t \equiv F(X_t)$ be the cumulative density function of state variables X_t. Let $\{X_t\}$ be a Markov process that the conditional density function of X_{t+1} given the information set $I_t \equiv \{X_t, X_{t-1}, \cdots\}$ only depends on its previous lagged variable X_t. We can rewrite the Euler equation (2.1) under the Markov assumption on $\{X_t\}$ as follows:

$$y_{t+1} = g_0(X_t, X_{t+1}) + \varepsilon_{t+1}, \tag{2.5}$$

where $y_{t+1} = m(X_{t+1})$, $g_0(X_t, X_{t+1}) = f(X_t) - m(X_{t+1})f(X_{t+1})$, and ε_{t+1} is an unobservable martingale difference sequence with respect to the information set I_t, namely $E(\varepsilon_{t+1}|I_t) = 0$, which follows from the Euler equation in (2.1).

Assumption 2.2.1. *The state variables X_t follows a Markov process and has a positive density function that is continuous almost everywhere on* $\mathbf{X}$.

In Equation (3.5), $\{\varepsilon_{t+1}\}$ can be interpreted as a sequence of aggregate pricing shocks. The martingale difference sequence property of $\{\varepsilon_{t+1}\}$ is a sufficient and necessary condition which guarantees the equivalence between the nonlinear time series regression model (3.5) and the Euler equation (2.1).

We shall estimate the unknown function $f^o(\cdot)$ by a global series approximation rather than local approximation techniques. Local constant and local polynomial approximations are among the most commonly used local estimation methods in nonparametric analysis (e.g., Fan and Gijbels, 1996). These methods are based on local Taylor expansions at a specific point. In Equation (3.5), at each time t, both $f(X_t)$ and $f(X_{t+1})$ contained in the regression function $g_0(X_t, X_{t+1})$ are unknown and must be estimated simultaneously. The difference $X_{t+1} - X_t$ varies with time, and its value can be large. Therefore, it is difficult, if not impossible, to pin down a common point suitable for the local Taylor expansions of both $f(X_t)$ and $f(X_{t+1})$. By contrast, series approximation provides a global solution to $f(\cdot)$ without seeking for local centers. By choosing a truncation order $p \equiv p(T) \to \infty$ as the sample size $T \to \infty$, the truncated series approximations to $f(X_t)$ and $f(X_{t+1})$ share a common set of coefficients in the entire domain $\mathbf{X}$ to be estimated from data. Therefore, a global series approximation avoids the problem that local approximation methods would encounter, and thus is the most convenient approach to estimating $f^o(x)$ for all $x \in \mathbf{X}$ in the present context.

To assist the consistency analysis of series estimation, we introduce some notations. The Sobolev norm is widely used in the nonparametric series regression literature. Andrews (1991) studies the asymptotic properties of nonparametric series estimation using the Sobolev norm. Newey (1997) and Newey and Powell (2003) further explore the rate of convergence of series estimation using the Sobolev norm as well. Let $\lambda = (\lambda_1, \cdots, \lambda_r)'$ be a vector of nonnegative integers, where r is the dimension of x, and let $\{\varphi_j(x)\}_{j=1}^{\infty}$ be a sequence of complete orthogonal basis functions. We define

$$\begin{aligned}
&|\lambda| = \sum_{i=1}^{r} \lambda_i,\ D^{|\lambda|} f(x) = \frac{\partial^{|\lambda|}}{\partial x^{\lambda_1} \cdots \partial x^{\lambda_r}} f(x), \\
&|f|_d = \max_{\lambda \le d} \sup_{x \in X} |D^{|\lambda|} f(x)|,\ \kappa_0(p) \ge |\varphi_j(x)|_0 \, \forall p \ge 1, \text{ and} \qquad (2.6) \\
&\xi_d(p) = \max_{|\lambda| \le d} \sup_{x \in \mathbf{X}} ||\partial^{\lambda} \varphi^p(x)|| \text{ for a vector } \varphi^p(x) = [\varphi_1(x), \cdots, \varphi_p(x)]',
\end{aligned}$$

where $||B|| = \sqrt{trace(B'B)}$ is the Euclidean norm for a vector or matrix B. To estimate the function $f^o(x)$ in model (3.5), we use a series approximation $f_p(x) \equiv \sum_{j=1}^{p} \alpha_j \varphi_j(x) = \varphi^p(x)'\alpha^p$, where $\alpha^p = (\alpha_1, \cdots, \alpha_p)'$ and the order p must grow to infinity as the sample size $T \to \infty$. We impose the following conditions on p and the basis functions.

Assumption 2.2.2. *Let $\{\varphi_j(x)\}_{j=1}^{\infty}$ be complete orthogonal basis functions. Suppose $f \in L^2$ is differentiable up to order $d \ge 0$. Let a truncated series $f_p(x) = \sum_{j=1}^{p} \alpha_j \varphi_j(x)$, where $p \equiv p(T) \to \infty$ as $T \to \infty$ such that (i) for an integer $d \ge 0$, there are $s > 0$ and $\alpha^p = (\alpha_1, \cdots, \alpha_p)'$ so that $|f^o - f_p|_d = O(p^{-s})$; (ii) $\frac{p^2}{T} \to 0$, $\frac{T}{p^{2s}} \to 0$ and $\frac{\xi_0^2(p)p}{T} \to 0$.*

Assumption 3.3.3 (i) is a rate condition at which the approximation bias $f^o(x) - f_p(x) = \sum_{j=p+1}^{\infty} \alpha_j \varphi_j(x)$ vanishes to zero as $p \to \infty$. Assumption 3.3.3 (ii) imposes conditions on p which imply that p must grow slower than $\sqrt{T}$ but faster than $T^{\frac{1}{2s}}$, so as to control both the variance and bias of the series estimator of $f^o(x)$.

Given different properties of the unknown function $f^o(x)$, we need to choose basis functions appropriately. We first enumerate some possible scenarios when $\mathbf{X}$ is compact. Suppose f^o is a periodic function over a compact support, say $\bar{Q} = [0, 2\pi]^l$, and is continuously differentiable up to order $r \in N$. Then we can choose the trigonometric series on $\bar{Q}$. Specifically, we consider the following Fourier series approximation:

$$f_p(x) = d_0 + \sum_{i=1}^{I_n} \sum_{j=1}^{J_n} \{d_{ij} cos(jk_i'x) + w_{ij} sin(jk_i'x)\} = \sum_{j=1}^{p} \alpha_j \varphi_j(x), \tag{2.7}$$

where $I_n,\ J_n \in N,\ d_0, d_{ij}$ and $w_{ij} \in R$. $k_i \in K_T \equiv \{k_i : i = 1, \cdots, I_n\}$ is an elementary multi-index, a $l \times 1$ vector of integers. For the construction of k_i, see Gallant (1981).

The periodicity assumption on f^o over a compact support appears strong. Relaxing it will result in boundary effects. Gallant and Souza (1991) introduce a Flexible Fourier Form (FFF) series to effectively improve the performance near the boundary. Hong and White (1995) further apply it to a nonparametric testing framework. Given the appealing advantages of the FFF series near the boundary, we consider a FFF series over $Q = [\nu, 2\pi - \nu]^l$ for any small $\nu > 0$:

$$f_p(x) = d_0 + \sum_{i=1}^{l} b_r x_i + \sum_{i=1}^{l} \sum_{j=1}^{i} c_{ij} x_i x_j + \sum_{i=1}^{I_n} \sum_{j=1}^{J_n} \{d_{ij} cos(jk_i'x) + w_{ij} sin(jk_i'x)\} = \sum_{j=1}^{p} \alpha_j \varphi_j(x), \tag{2.8}$$

where $(\alpha_1, \cdots, \alpha_p) = (d_0, \alpha_{(0)}, \alpha_{(1)}, \cdots, \alpha_{(I_n)})$, where $\alpha_{(0)} = (b_1, \cdots, b_d, c_{11}, c_{12}, \cdots, c_{dd})$ and $\beta_{(i)} = (d_{i1}, w_{i1}, \cdots, d_{iJ_n}, w_{iJ_n})$. For the construction of a FFF series, see Gallant and Souza (1991).

Regression splines are also popular in approximating unknown functions with compact support. Let $Q = [0,1]^l$ and $\Delta = \{s_i\}_{i=1}^{k}$ with $0 = s_1 < s_2 \cdots < s_{k+1} = 1$ be a partition of Q into q intervals

$$I_i = [s_i, s_{i+1}),\ i = 1, \cdots, q-1 \,\text{and}\, I_k = [s_k, s_{k+1}]. \tag{2.9}$$

Suppose $f(x)$ is a w-th order polynomial $\in C^{w-2}$ at $s_i, i = 1, \cdots, k$, where $w \in N$. The space of polynomial splines of order w with knots $s_1, \cdots, s_k$ is defined as

$$f_p(x) = \sum_{j=1}^{p} \alpha_j \varphi_j(x) \,\text{for}\, x \in I_j. \tag{2.10}$$

A direct choice for $\{\varphi_j(x)\}_{j=1}^{\infty}$ is the normalized w-th order B-splines $\{N_j^w\}$ with knots $s_j, \cdots, s_{j+w}$ that satisfy $\sum_{i=j+1-w}^{j} N_j^w(x) = 1$ for all $s_j \leq x < s_{j+1}$.

Finally, we consider the situation when $\mathbf{X}$ is unbounded. In this case, we consider convergence in the weighted supremum norm $|f|_{\infty,w} \equiv \sup_{x\in\mathbf{X}} |f(x)w(x)|$, where $w : R \to R$ is a weighting function. When $w(x) = 1$, we have $|f|_0 = |f|_{\infty,w(x)=1}$, as a special case of the weighted supremum norm. We consider a Hermite series approximation:

$$f_p(x) = \sum_{j=1}^{p} \alpha_j \varphi_j(x), \tag{2.11}$$

where $\varphi_j(x) = w(x)H_j(x)$, $H_j(t) = (-1)^j e^{t^2} \frac{d^j e^{-t^2}}{dt^j}$, and a weighting function $w(x) = e^{-x^2}$.

Given the choice of basis functions $\{\varphi_j(x)\}_{j=1}^{\infty}$, we define

$$\psi_{j,t} \equiv \varphi_j(X_t) - m(X_{t+1})\varphi_j(X_{t+1}) \,\text{for all}\, t = 1, 2, \cdots, T \,\text{and}\, j = 1, 2, \cdots. \tag{2.12}$$

Then the nonlinear regression model (3.5) can be represented equivalently as the following generalized linear regression model:

$$y_{t+1} = \sum_{j=1}^{\infty} \alpha_j \psi_{j,t} + \varepsilon_{t+1},\ t = 1, 2, \cdots T. \tag{2.13}$$

Given Assumption 3.3.3, we can use the following truncated series regression

$$y_{t+1} = \sum_{j=1}^{p} \alpha_j \psi_{j,t} + u_{p,t+1},\ t = 1, 2, \cdots T, \tag{2.14}$$

where $u_{p,t+1} = \varepsilon_{t+1} + \sum_{j=p+1}^{\infty} \alpha_j \psi_{j,t}$, and the bias term $\sum_{j=p+1}^{\infty} \alpha_j \psi_{j,t}$ vanishes to zero in probability provided $p \to \infty$ sufficiently fast as $T \to \infty$.

In conventional regression analysis, endogeneity is mainly caused by omitted variables, measurement errors, simultaneous equation biases, and peer effects. Here, endogeneity arises as a result of recursive occurrences of the unknown function $f^o(\cdot)$ over two time periods. The control variable $\psi_{j,t}$ contains an ingredient $m(X_{t+1})$ which leads to correlation between the control variable and the true regression error, namely $E(\psi_{j,t}\varepsilon_{t+1}) \neq 0$ for at least one $j \in \{1, 2, \cdots\}$. As a consequence, the OLS series estimation will not be consistent for Equation (2.13). For consistent estimation of the price-dividend ratio function $f^o(x)$, we introduce instrumental variables (IV) to eliminate endogeneity biases. Suppose Z_t is an instrumental vector $Z_t \in \mathbf{X}$ so that

$$E(\varepsilon_{t+1}\Phi_{q,t}) = \mathbf{0}, \text{ and } E(\psi_{j,t}\Phi_{q,t}) \neq \mathbf{0} \text{ for } j = 1, 2, \cdots, q;\ t = 1, 2, \cdots, T, \tag{2.15}$$

where $\Phi_{q,t} = [\phi_1(Z_t), \cdots, \phi_q(Z_t)]'$ for some basis functions $\{\phi_j(z)\}_{j=1}^{q}$ which may differ from the basis functions $\{\varphi_j(x)\}_{j=1}^{p}$. An example of Z_t is to choose $Z_t = X_t$.

Then $E(\varepsilon_{t+1}|X_t) = 0$ implies $E(\varepsilon_{t+1}|\Phi_{q,t}) = \mathbf{0}$. The second condition in Equation (2.15) usually holds because the set of instrumental variables $\Phi_{q,t}$ is correlated with control variables $\{\psi_{j,t}\}_{j=1}^{p}$.

To ensure identification of coefficients $\alpha^p = (\alpha_1, \cdots, \alpha_p)'$ in Equation (2.14), we must have $q \geq p$. Denote a $T \times p$ control variable matrix $\Psi_{Tp} = (\Psi_{p,1}, \cdots, \Psi_{p,T})'$ with $\Psi_{p,t} = (\psi_{1,t}, \cdots, \psi_{p,t})'$, a $T \times q$ instrumental variable matrix $\Phi_{Tq} = (\Phi_{q,1}, \cdots, \Phi_{q,T})'$ with $\Phi_{q,t} = (\phi_{1,t}, \cdots, \phi_{q,t})'$, $Y = (y_1, \cdots, y_T)'$, and $\varepsilon = (\varepsilon_1, \cdots, \varepsilon_T)'$. Then consistent estimation of the function f^o can be obtained using the following 2SLS series regression. Without abuse of notations, we suppress subscripts in Ψ_{Tp} and Φ_{Tq}, and let $q = p$ for simplicity.

In the first stage, we regress control variables $\Psi_{p,t}$ on instrumental variables $\Phi_{q,t}$ to obtain fitted values $\hat{\Psi}_{p,t}$ via OLS. Specifically, for each endogenous control variable $\psi_{j,t}$, $1 \leq j \leq p$, we run a simple auxiliary regression model

$$\psi_{j,t} = \Phi'_{q,t} a_j + v_t, \text{ for } t = 1, \cdots, T. \tag{2.16}$$

The fitted values are $\hat{\Psi} = \Phi(\Phi'\Phi)^{-1}\Phi'\Psi$, with $\hat{\Psi}'_{p,t}$ being the t-th row of $\hat{\Psi}$.

In the second stage, we regress y_{t+1} on regressors $\hat{\Psi}_{p,t}$ to estimate the coefficients $\{\alpha\}_{j=1}^{p}$ in the following model

$$y_{t+1} = \hat{\Psi}'_{p,t}\alpha^p + \tilde{u}_{t+1}, \text{ for } t = 1, 2, \cdots, T. \tag{2.17}$$

With the OLS estimator $\hat{\alpha}^p = (\hat{\Psi}'\hat{\Psi})^{-1}\hat{\Psi}'Y$, we obtain a closed-form estimator

$$\hat{f}_p(x) = \sum_{j=1}^{p} \varphi_j(x)\hat{\alpha}_j = \varphi^p(x)'(\hat{\Psi}'\hat{\Psi})^{-1}\hat{\Psi}'Y = \varphi^p(x)'[\Psi'\Phi(\Phi'\Phi)^{-1}\Phi'\Psi]^{-1}[\Psi'\Phi(\Phi'\Phi)^{-1}\Phi'Y]. \tag{2.18}$$

One appealing feature of this 2SLS procedure is its easy implementation. It always has a data-based closed-form solution, no matter how complex the DSGE model is. Also, we do not have to specify the DGP for state variables $\{X_t\}$. While estimating $f^o(x)$ by the generalized method of moments (GMM) can generally enhance estimation efficiency, the computational cost of deriving the optimal weighting matrix in GMM is a concern. Although Gao and Hong (2015) suggest a new Bayesian resampling method, the implementation of the optimal GMM estimation is still a thorny procedure and usually does not have a closed-form solution. Using the convenient 2SLS series regression procedure, we can consistently estimate the unknown function $f^o(x)$ without involving any numerical integration or optimization. Furthermore, if the DSGE model has conditional homoskedastic pricing shocks, namely $E(\varepsilon_{t+1}^2|\Phi_{q,t}) = \sigma^2$ for all t, then the 2SLS series estimator is asymptotically efficient.

2.3 Consistency and Asymptotic Normality

The 2SLS series regression procedure differs from the classical 2SLS procedure in that the number of regressors in both stages grows to infinity as the sample size $T \to \infty$, and there exists a bias term due to the finite order series approximation $f_p(x)$ for the unknown function $f^o(x)$. We first show that the 2SLS series regression procedure yields a consistent estimator for $f^o(x)$. For this purpose, we impose some mild conditions on state variables X_t and the unobservable aggregate pricing shock ε_{t+1}.

Assumption 2.3.1. *For all $t \geq 0$, some $\delta > 0$ and $0 < \Delta < \infty$ (i) $\{X_t, \varepsilon_{t+1}\}$ is an $\alpha-$mixing sequence with mixing coefficients $\alpha(j)$ so that $\sum_{j=1}^{\infty} \alpha(j)^{\frac{\delta}{4+\delta}} < \Delta$; (ii) $E|\phi_{j,t}|^{4+\delta} < \Delta$, $j = 1, \cdots, p$, and $E|\varepsilon_{t+1}|^{4+\delta} < \Delta$; (iii) $E|\varphi_{j,t}|^{8+\delta} < \Delta$, $j = 1, \cdots, p$, and $E|m(X_{t+1})|^{8+\delta} < \Delta$.*

Let $\lambda_{\min}[E(\frac{\Phi'\Phi}{T})]$ and $\lambda_{\max}[E(\frac{\Phi'\Phi}{T})]$ denote the minimum and maximum eigenvalues of a $p \times p$ matrix $E(\frac{\Phi'\Phi}{T})$ respectively, where $p \to \infty$ as $T \to \infty$. We impose some mild conditions on $\lambda_{\min}[E(\frac{\Phi'\Phi}{T})]$ and $\lambda_{\max}[E(\frac{\Phi'\Phi}{T})]$ so that consistent estimation of the parameters α^p in the 2SLS series regression can be obtained.

Assumption 2.3.2. *For all $p \geq 1$ (i) $\lambda_{\min}[E(\frac{\Phi'\Phi}{T})] > 0$; (ii) $\lambda_{\max}[E(\frac{\Phi'\Phi}{T})] < \infty$; (iii) $\lambda_{\max}[E(\frac{\Phi'\varepsilon\varepsilon'\Phi}{T})] < \infty$.*

Assumption 3.3.5 (i) is the well-known necessary and sufficient condition for consistent estimation of parameters in a linear regression model with a fixed number of regressors (Drygas, 1976). This assumption is also employed by Andrews (1991) to establish consistency of the OLS series estimator when the number of regressors grows to infinity as $T \to \infty$. With a stronger assumption that $\lambda_{\min}[E(\frac{\Phi'\Phi}{T})]$ is uniformly bounded away from below from zero, Portnoy (1985) imposes Assumption 3.3.5 (ii) to obtain consistent estimation of parameters in a general linear regression model when the number of regressors tends to infinity as $T \to \infty$. As pointed out by Andrews (1991), Assumption 3.3.5 (i) holds with probability one if $E(\Phi_{p,t}\Phi'_{p,t})$ is nonsingular for all $p \geq 1$. When this condition is violated, basis functions that are redundant in the limit can be eliminated to ensure that Assumption 3.3.5 always holds.

Suppose the density dF_t/dX_t of state variables X_t with a compact support is bounded away from below from zero. Then it is easy to prove that $\lambda_{\min}E(\frac{\Phi'\Phi}{T})$ is bounded away from below from zero uniformly, namely $\lambda_{\min}E(\frac{\Phi'\Phi}{T}) > c > 0$ for some constant $c > 0$. However, we do not require that the minimum eigenvalue of $E(\frac{\Phi'\Phi}{T})$ be uniformly bounded away from below from zero for all circumstances. For instance, when we use FFF to estimate a non-periodic function $f^o(x)$, as shown by Gallant and Souza (1991), $\lambda_{\min}E(\frac{\Phi'\Phi}{T})$ decreases to zero rapidly. More specifically, in this case $\lambda_{\min}E(\frac{\Phi'\Phi}{T}) = O(p^{-(s+\epsilon)/l}) \to 0$ for every $s \in N$ and $\epsilon > 0$. Also, as pointed out by Hong and White (1995), the choice of normalized B-splines leads to $\lambda_{\min}E(\frac{\Phi'\Phi}{T}) = O(p^{-1}) \to 0$.

In fact, the convergence rate of $\lambda_{\min}E(\frac{\Phi'\Phi}{T})$ is a key ingredient that affects the rate at which $p \to \infty$ as $T \to \infty$. Hong and White (1995) establish a series of insightful results on the rate of p in order to to ensure that their test statistics have a well-behaved asymptotic distribution when $\lambda_{\min}E(\frac{\Phi'\Phi}{T})$ is not uniformly bounded from below from zero. Andrews (1991) considers how the minimum eigenvalue can affect consistency of a series estimator. To cover various cases, we define a positive function $\lambda : N_+ \to [1, \infty)$ so that

$$\lambda(p)\lambda_{\min}[E(\frac{\Phi'\Phi}{T})] = 1 \,\text{for all}\, p \geq 1. \tag{2.19}$$

Intuitively, $\lambda(p)$ is the inverse $\lambda_{\min}[E(\frac{\Phi'\Phi}{T})]$. Relaxing the condition on $\lambda_{\min}[E(\frac{\Phi'\Phi}{T})]$ requires a careful analysis on the behavior of its sample analog $\frac{\Phi'\Phi}{T}$.

Assumption 2.3.3. $\frac{p\lambda^2(p)}{\sqrt{T}} \to 0$ *and* $\frac{p^2\lambda(p)}{\sqrt{T}} \to 0$.

Because both $E(\frac{\Phi'\Phi}{T})$ and its sample analog $\frac{\Phi'\Phi}{T}$ are $p \times p$ matrices where the dimension $p \to \infty$, Assumption 3.3.6 imposes some restrictions on the rate of $\lambda(p)$ to ensure that $\lambda_{\min}(\frac{\Phi'\Phi}{T})$ of the sample matrix $\frac{\Phi'\Phi}{T}$ converges to $\lambda_{\min}[E(\frac{\Phi'\Phi}{T})]$ almost surely. This is essential for the consistency of the proposed 2SLS series estimator. We now establish consistency of the series estimator $\hat{f}_p(x)$ in Theorem 3.3.4 below.

Theorem 2.3.1 (Consistency)**.** *Suppose Assumptions 3.2.3-3.3.6 hold. Then there exists a unique solution $f^o(x)$ to Equation (2.1), and the nonparametric 2SLS series estimator $\hat{f}_p(x)$ satisfies:*

(i) $\int [\hat{f}_p(x) - f^o(x)]^2 dF(x) = O_p[\lambda(p)(\frac{p}{T} + p^{-2s})]$;

(ii) with $d \geq 0$ and any given $x \in \mathbf{X}$, $|\hat{f}_p(x) - f^o(x)|_d = O_p[\xi_d(p)\sqrt{\lambda(p)(\frac{p}{T} + p^{-s})}]$;

(iii) with $d \geq 0$, $p > lnT$, $\frac{\xi_d(p)^2\lambda(p)ln(T)}{T} \to 0$ and any given $x \in \mathbf{X}$, $|\hat{f}_p(x) - f^o(x)|_d = O_P[\xi_d(p)\sqrt{\lambda(p)}(\sqrt{\frac{lnp}{T}} + (1 + |f_p|_{\infty,w})p^{-s}]$.

Theorem 3.3.4 (i) is a global consistency result, while Theorem 3.3.4 (ii, iii) are pointwise consistency results. For example, when using splines to estimate $f^o(x)$, we have $|f_p|_{\infty,w} = O(1)$ as proved in Huang (2003), and Theorem 3.3.4 (iii) implies $|\hat{f}_p(x) - f^o(x)|_d = O_P[\xi_d(p)\sqrt{p}(\sqrt{\frac{lnp}{T}} + p^{-s})]$. Compared with the existing numerical solution methods, our procedure has a convenient data-based closed-form solution regardless the complexity of the DSGE model. Most importantly, Theorem 3.3.4 implies that our procedure is always free of misspecification for the price-dividend ratio function when the sample size $T \to \infty$,and we do not have to specify the DGP for state variables. This appealing property is not attainable by existing numerical solution methods in the literature that have to specify a model for the DGP of state variables, which therefore may suffer from model misspecification.

We emphasize that our series regression differs from the projection method in the literature in several dimensions. First, we require the order p to grow to infinity at a suitable rate as the sample size $T \to \infty$. Second, we do not have to specify the DGP of state variables and estimate model parameters in the DGP. Third, we estimate the projection coefficients from data, rather than solving numerical integrations. Finally, the projection method generally does not have a closed-form solution, especially when the CAPM becomes complex.

The convergence result obtained in Theorem 3.3.4 is completely independent of models and parameter values such as the curvature of the utility function and the time discount rate. Previously, solution accuracy was evaluated by comparing an approximation solution with an analytic one (e.g., Tauchen and Hussey, 1991 and Christiano, 1990). Concerned with the availability of an analytic solution, Den Haan and Marcet (1994) propose a χ^2 test for the accuracy of approximation solutions. An issue with this testing method is that orthogonal Euler equation residuals may be compatible with large deviations from the optimal policy (Santos, 2000), thereby not being effective in all circumstances. By examing the curvature of the utility function and other related parameter values, Santos (2000) show that the approximation errors of these unknown functions are of the same order of magnitude as the size of the Euler equation residuals. Despite these improvements in quantifying the accuracy of an approximation solution, no direct result on the convergence rate between an approximated function and its exact solution has been reached. Using the nonparametric 2SLS series regression procedure, this paper derives a closed-form result on the rate that approximation errors and sampling variance vanish to zero as $T \to \infty$ without seeking additional help from the size of Euler equation errors or comparison with solutions using extremely fined grids.

Theorem 3.3.4 provides a range of admissible rates for p. In practice, one may like to choose p via data-driven methods, such as using the AIC or BIC criterion. We will investigate this issue in our empirically relevant simulation studies.

To make rigorous statistical inference such as confidence interval estimation and hypothesis testing, we shall derive the asymptotic distribution of the series estimator $\hat{f}_p(x)$. Put $S_{pT} = E(\frac{\Phi'\varepsilon\varepsilon'\Phi}{T})$, $Q_T = E(\frac{\Phi'\Psi}{T})$ and $P_T = E(\frac{\Phi'\Phi}{T})$, all of which are $p \times p$ matrices. Define $V_{pT} \equiv \varphi^p(x)'E(\frac{\Phi\varepsilon\varepsilon'\Phi'}{T})\varphi^p(x) = \varphi^p(x)'S_{pT}\varphi^p(x)$. Then the variance of the series estimator $\hat{f}_p(x)$ is

$$D_{pT}(x) = \varphi^p(x)'(Q_T'P_T^{-1}Q_T)^{-1}Q_T'P_T^{-1}S_{pT}P_T^{-1}Q_T(Q_T'P_T^{-1}Q_T)^{-1}\varphi^p(x). \tag{2.20}$$

If there exists conditional homoskedasticity, i.e., $E(\varepsilon_{t+1}^2|\Phi_{q,t}) = \sigma^2$ for all t, then we have

$$D_{pT}(x) = \sigma^2\varphi^p(x)'(Q_T'P_T^{-1}Q_T)^{-1}\varphi^p(x). \tag{2.21}$$

Assumption 2.3.4. *For some $\delta > 0$, $p \to \infty$ as $T \to \infty$, (i) $E|\varepsilon_t|^{8+\delta} < \Delta < \infty$ for all t; (ii) $\frac{\lambda^5(p)p\xi_0(p)}{\sqrt{T}} \to 0$; (iii) $\frac{\lambda^4(p)p^3}{T} \to 0$.*

Assumption 2.3.4 provides a set of sufficient conditions on unobservable pricing shocks, the rates of order p and $\lambda(p)$ so that the following asymptotic normality can be established.

Theorem 2.3.2 (Asymptotic Normality)**.** *Suppose Assumptions 3.2.3-2.3.4 hold. Then for any given $x \in \mathbf{X}$, as $T \to \infty$,*

(i)

$$\sqrt{\frac{T}{D_{pT}(x)}}[\hat{f}_p(x) - E\hat{f}_p(x)] \xrightarrow{d} N(0,1); \tag{2.22}$$

(ii)

$$\sqrt{\frac{T}{D_{pT}(x)}}[\hat{f}_p(x) - f^o(x)] \xrightarrow{d} N(0,1). \tag{2.23}$$

Theorem 3.3.5 (i) and (ii) imply that the bias of the series estimator $\hat{f}_p(x)$ vanishes to zero sufficiently fast so that it does not affect the asymptotic normal distribution of $\hat{f}_p(x)$.

Our method is also applicable to hidden Markov processes. Suppose state variables X_t is not directly observable, but can be estimated via such methods as Kalman filters.

Theorem 2.3.3 (Hidden Markov Processes)**.** *Suppose Assumptions 3.2.3-2.3.4 hold, and $\hat{x}$ is a $\sqrt{T}$-consistent estimator for some given point $x \in \mathbf{X}$. Then as $T \to \infty$,*

(i)

$$\sqrt{\frac{T}{D_{pT}(\hat{x})}}[\hat{f}_p(\hat{x}) - E\hat{f}_p(x)] \xrightarrow{d} N(0,1); \tag{2.24}$$

(ii)

$$\sqrt{\frac{T}{D_{pT}(\hat{x})}}[\hat{f}_p(\hat{x}) - f^o(x)] \xrightarrow{d} N(0,1). \tag{2.25}$$

Intuitively, the estimated state variables $\hat{x}$ converges in probability to the point x at a parametric rate $T^{-\frac{1}{2}}$, which is faster than the convergence rate of the non-parametric series estimator $\hat{f}_p(x)$ to $f^o(x)$. As a result, the sampling errors of the estimator $\hat{x}$ of x do not have impact on the asymptotic distribution of $\hat{f}_p(\hat{x})$.

We now consider estimation of the variance D_{pT} of the series estimator $\hat{f}_p(x)$ given $x \in X$. To allow for conditional heteroskedasticity of ε_{t+1}, we define a $p \times p$ matrix

$$\hat{S}_{pT} = \frac{1}{T}\sum_{t=1}^{T} \Phi_{p,t}\Phi'_{p,t}\hat{\varepsilon}^2_{t+1}, \tag{2.26}$$

where $\hat{\varepsilon}_{t+1} = y_{t+1} - \sum_{i=1}^{p} \psi_{i,t}\hat{\alpha}_i$. Note that $\hat{\varepsilon}_{t+1}$ is not the estimated residual of the second stage regression. Then we obtain a heteroskedasticity-robust variance estimator for $\hat{f}_p(x)$:

$$\hat{D}_{pT}(x) = \varphi^p(x)'[\frac{\Psi'\Phi}{T}(\frac{\Phi'\Phi}{T})^{-1}\frac{\Phi'\Psi}{T}]^{-1}(\frac{\Phi'\Psi}{T})(\frac{\Phi'\Phi}{T})^{-1}\hat{S}_{pT}(\frac{\Phi'\Phi}{T})^{-1}(\frac{\Psi'\Phi}{T})[\frac{\Psi'\Phi}{T}(\frac{\Phi'\Phi}{T})^{-1}\frac{\Phi'\Psi}{T}]^{-1}\varphi^p(x). \tag{2.27}$$

Assumption 2.3.5. *With $p \to \infty$ as $T \to \infty$, (i) $\frac{\lambda^7(p)p}{\sqrt{T}} \to 0$; (ii) $\sqrt{\lambda(p)}(\sqrt{\frac{p}{T}} + p^{-s})p^2 \to 0$.*

For any given $x \in \mathbf{X}$, Assumption 2.3.5 (i) ensures that the ratio of the variance estimator $\hat{D}_{pT}(x)$ to the population variance $D_{pT}(x)$ converges to 1 in probability, because we can show $\frac{\hat{D}_{pT}(x)}{D_{pT}(x)} - 1 = O_p(\frac{\lambda^7(p)p}{\sqrt{T}}) = o_p(1)$. Moreover, Assumption 2.3.5 (ii) guarantees that the sampling errors and approximation errors do not affect the consistency of $\hat{D}_{pT}(x)$.

Theorem 2.3.4 (Consistent Variance Estimation)**.** *Suppose Assumptions 3.2.3-2.3.5 hold. Then for any given $x \in \mathbf{X}$, as $T \to \infty$,*

(i) $\frac{\hat{D}_{pT}(x)}{D_{pT}(x)} \xrightarrow{p} 1$;

(ii) $\sqrt{\frac{T}{\hat{D}_{pT}(x)}}[\hat{f}_p(x) - f^o(x)] \xrightarrow{d} N(0,1)$.

It is worth mentioning that the consistently estimated price-dividend ratio function can be further applied to constructing correct inference on equity premiums, and conducting valid model evaluation. The estimated return of the risky asset R_{t+1} in each time period $t+1$ can be uniquely expressed as

$$\hat{R}_{t+1} = \frac{\hat{f}_p(X_{t+1}) + 1}{\hat{f}_p(X_t)} e^{X_{t+1}}. \tag{2.28}$$

The 2SLS series regression method provides direct channel to compute the estimated risky returns via $\hat{f}_p$. Therefore, for each set of model structural parameters κ, we can consider the simulated moments as follows:

$$\hat{m}(\kappa) = \begin{cases} \frac{1}{T}\sum_{t=1}^{T-1}\hat{R}_{t+1} - \frac{1}{T}\sum_{t=1}^{T-1}R_{t+1} \\ \frac{1}{T}\sum_{t=1}^{T-1}[\hat{R}_{t+1} - \frac{1}{T}\sum_{t=1}^{T-1}\hat{R}_{t+1}]^2 - \frac{1}{T}\sum_{t=1}^{T-1}[\hat{R}_{t+1} - \frac{1}{T}\sum_{t=1}^{T-1}R_{t+1}]^2 \end{cases} \tag{2.29}$$

where R_{t+1} is the empirically observed real equity returns.

This set of moment conditions can be included in the SMM procedure, providing a direct link between DSGE models and real data without involving any estimation biases due to approximation errors of the price-dividend ratio function and the DGP of state variables. The 2SLS series regression approach provides an additional dimension for controlling the explanatory ability of model parameters by including both Euler equation based moments and sample moments when fitting microeconomic data.

2.4 Simulation Studies and Empirical Applications

2.4.1 The Mehra and Prescott (1985) Model

The Case with a Correctly Specified DGP

We now compare the finite sample performance of the nonparametric 2SLS series regression procedure with popular numerical solution methods in solving the famous Mehra and Prescott's (1985) CAPM. An infinitely-lived representative agent wishes to maximize her expected lifetime utility at time zero

$$\begin{aligned} &\max_{\{C_t\}} E \sum_{t=0}^{\infty} \beta^{t-1} \frac{C_t^{1-\gamma}}{1-\gamma} \\ &\text{s.t.}\, C_t + P_{t+1}\theta_{t+1} + Q_t b_{t+1} = b_t + (D_t + P_t)\theta_t, \end{aligned} \tag{2.30}$$

where $X_t = ln(C_t/C_{t-1})$, $X_{t+1} - \mu = \Gamma(X_t - \mu) + u_{t+1}$, and $u_{t+1} \sim i.i.d.N(0, \sigma_u^2)$.

In this simple economy, dividend payment D_t is equal to consumption C_t in equilibrium. The Euler equation can be derived as follows:

$$f_t = \beta E[e^{(1-\gamma)X_{t+1}}(f_{t+1}+1)|X_t]. \tag{2.31}$$

We use normalized Hermite polynomials as complete orthogonal basis functions $\{\varphi_j(x)\}_{j=1}^{\infty}$. We consider five sets of parametrizations and report our estimation results in comparison with those obtained from popularly used numerical solution approaches in Table 3.1. In Table 3.1, five GDPs are selected to include a broad range of situations and also guarantee the existence of analytic solutions, which provide the best benchmark to compare the goodness of different solution methods.

Table 2.1: DGP for Model 1 Case 1

	Parametrizations of preferences				
	β	γ	Γ	μ	σ_u
DGP 1	0.95	2.50	−0.139	1.79%	3.48%
DGP 2	0.95	2.50	0.139	1.79%	3.48%
DGP 3	0.95	2.50	0.8	1.79%	3.48%
DGP 4	0.95	2.50	−0.8	1.79%	3.48%
DGP 5	0.95	31	−0.139	1.79%	3.48%

Note: β is the time-discount factor, γ specifies the risk-aversion level, Γ controls the autocorrelation of the annual consumption growth rates, μ is the unconditional mean of the annual consumption growth rate and σ_u is the standard deviation of the consumption growth rates.

In Figures 1-5, we plot the estimated price-dividend ratios together with the approximation results obtained from the perturbation and projection methods under all five DGPs. Using suitable information criterion such as AIC, a proper order p of series approximation can be decided. Specifically, the projection method that we consider is the Garlerkin method. For a fair judgement, we also plot the analytic price-dividend ratios as a benchmark to visualize the goodness of fit because these five DGPs are chosen so that analytic solutions exist. First, we observe that the analytic solutions in DGPs 1 and 2, where low persistency in absolute values is assumed, look linear. Figures 1 and 2 illustrate that both the perturbation and projection methods can perfectly capture the linear function, which is not surprising because approximation errors are extremely weak in these cases. The 2SLS series regression procedure also performs very well in these two situations. The mean squared errors reported in Table 2.2 are larger than that from the other two numerical methods, which are apparently due to sampling variations in estimation. However, the estimate is rather close to the analytic solution and is securely within the 95% confidence interval band.

Figures 3 and 4 plot the results under DGPs 3 and 4, where high persistency in absolute values is assumed. The projection method becomes ineffective, falling far away from the 95% confidence interval band. Apart from its sensitivity to the choice of polynomial order p, another limitation of the projection method is that a high persistency rate Γ of the Markov process $\{X_t\}$ results in a loss of accuracy due to the wide boundary of state variables. In contrast, the 2SLS series regression is particularly suitable for state variables with high persistency, which significantly improves estimation efficiency given our choice of instruments $Z_t = X_t$ (so that $\Phi_{p,t}$ is highly correlated with $\Psi_{q,t}$; see Equation (2.15)). The perturbation method outper-

Figure 2.1: P/D under DGP 1

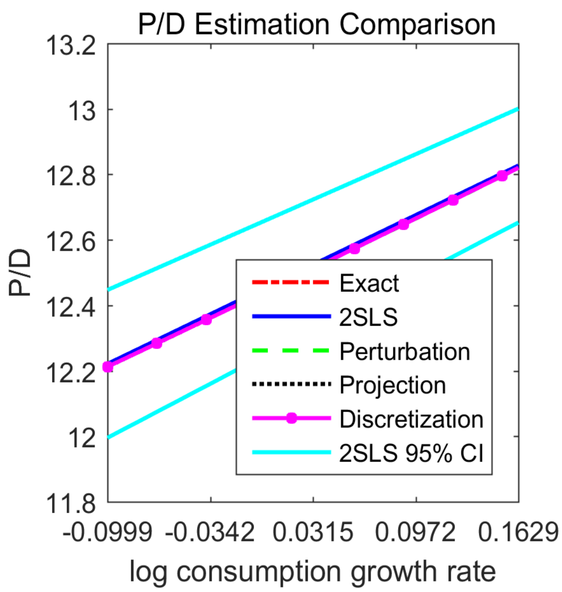
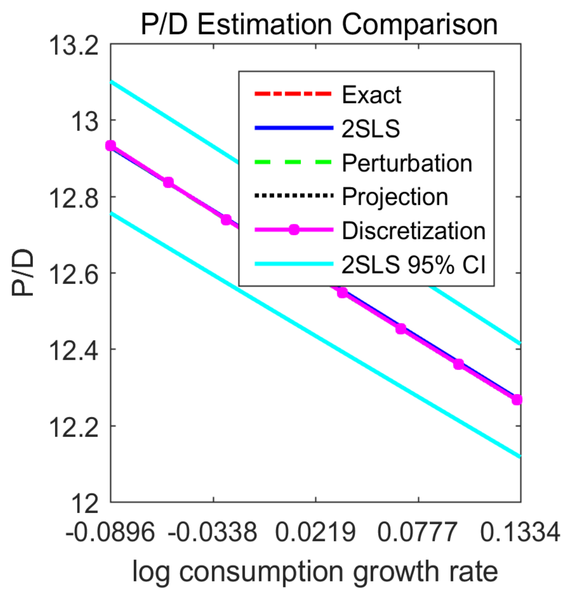


Figure 2.2: P/D under DGP 2



forms the projection method, but still cannot fit the exact solution well in two tails. The functional form misspecification of the perturbation method becomes substantial when the true price-dividend ratio function has a large curvature. Both the tail areas and the region around steady states are poorly approximated by misspecified solutions. In contrast, the 2SLS series regression outperforms both the projection and perturbation methods. Overall, our method is successful in fitting the exact solution in the entire support.

In Figure 5, the representative agent has a risk aversion level as high as $\gamma = 31$. This scenario is crucial given its theoretical importance. The exact solution becomes nonlinear. The perturbation method performs the worst. It fails to mimic the dynamics over the entire domain. The projection method provides a better approximation than the perturbation method, but the 2SLS series regression provides

Figure 2.3: P/D under DGP 3

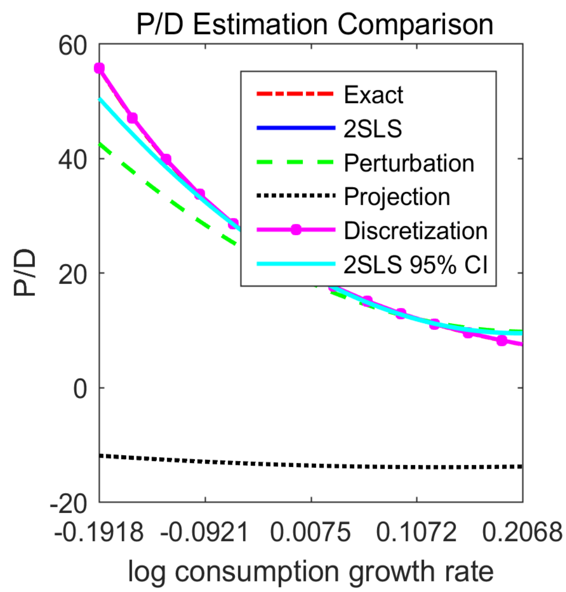
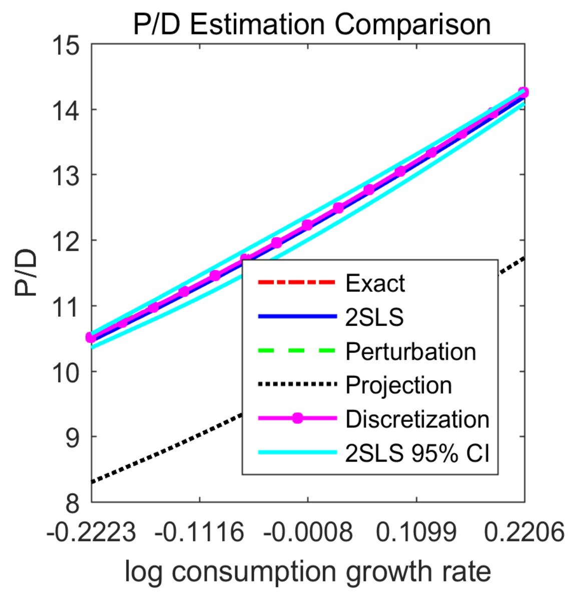


Figure 2.4: P/D under DGP 4



the best goodness of fit over the entire domain.

Generally speaking, the perturbation method suffers from approximation errors, especially when the true price-dividend ratio function is highly nonlinear and over tail distributions. The projection method is largely limited by the level of persistency, and also suffers from approximation errors that do not disappear as the sample size $T \to \infty$. The 2SLS series regression performs the best in estimating price-dividend ratios because of its reasonable and robust performance over various parametrizations. In fact, unlike the existing numerical solution methods, our procedure does not require any specification of the DGP for state variables $\{X_t\}$, and we allow a rather general class of stationary Markov processes for $\{X_t\}$.

A central concern of the asset pricing literature is understanding the well-known equity premium puzzle. Misspecification errors in approximating the CPAM contaminate the results on equity premiums in a non-negligible manner. Table 2.2 reports the first two moments of the risk-free asset and the equity premiums using the 2SLS series regression method, the perturbation method and the projection method respectively under the five DGPs. Analytic solutions are used as a benchmark under each set of comparisons. We also report the mean absolute difference (MAD_f) and the mean squared errors (MSE_f) between the analytic solution and each specific solution method as well as the mean absolute difference (MAD_{EL}) and the mean squared errors (MSE_{EL}) of the Euler equation. These four statistics can be used as a quick and sharp means for judging the goodness of various methods.

The first two moments of the risk-free asset are not affected by how price-dividend ratios are solved because the returns of the risk-free asset can be solved analytically. Compared with the projection and perturbation methods, the 2SLS series regression provides the most reasonable simulated first-two moments of asset returns under all situations. It overcomes the obstacles of the other two methods. In DGPs 3-5, where both the projection and perturbation methods fail, the 2SLS series regression delivers the most accurate results on equity premiums. Furthermore, it provides a reliable estimation of the correlation between equity premiums and risk-free returns even in the worst performing case. Finally, we compare the 2SLS series regression with the OLS series estimation. In Figures 6-10, for all five situations it is very clear that the 2SLS series regression significantly outperforms the OLS series estimation. This is because the latter suffers from endogeneity biases and thus is not consistent.

Figure 2.5: P/D under DGP 5

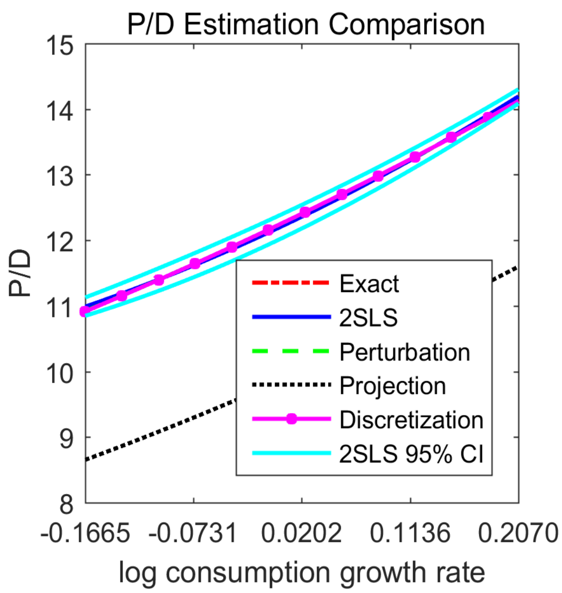


Figure 2.6: 2SLS Series vs OLS under DGP 1

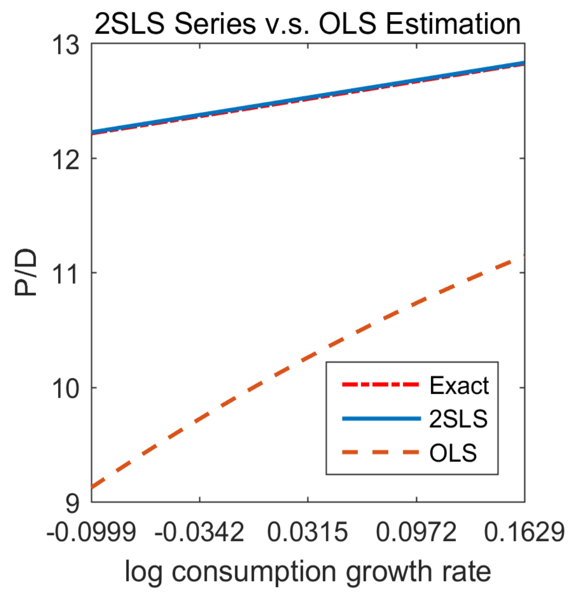


Figure 2.7: 2SLS Series vs OLS vs OLS under DGP 2

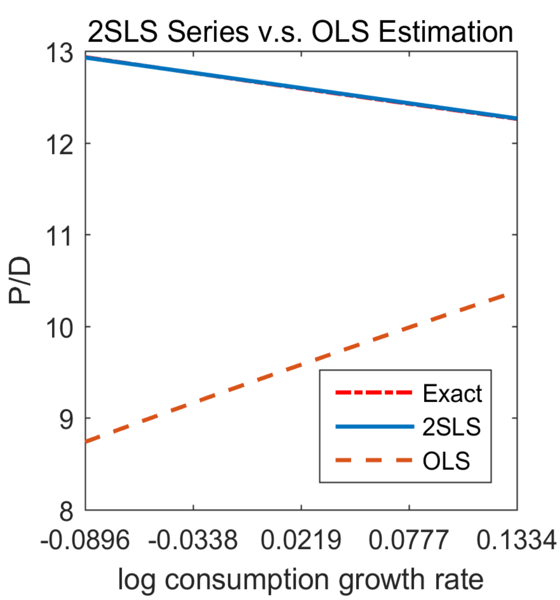


Figure 2.8: 2SLS Series vs OLS under DGP 3

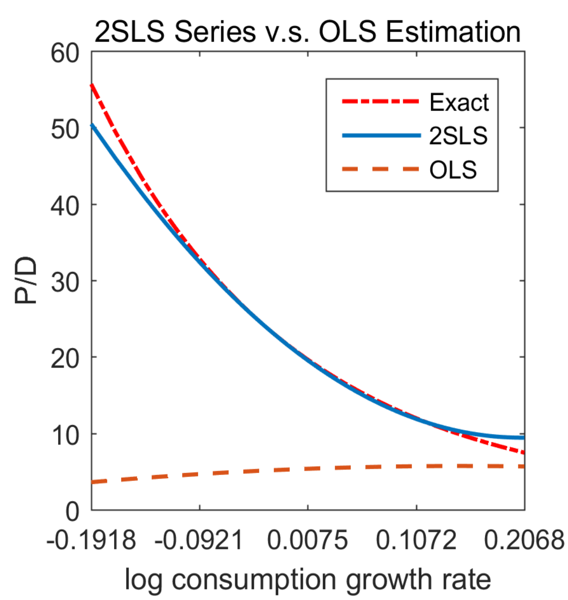


Figure 2.9: 2SLS Series vs OLS under DGP 4

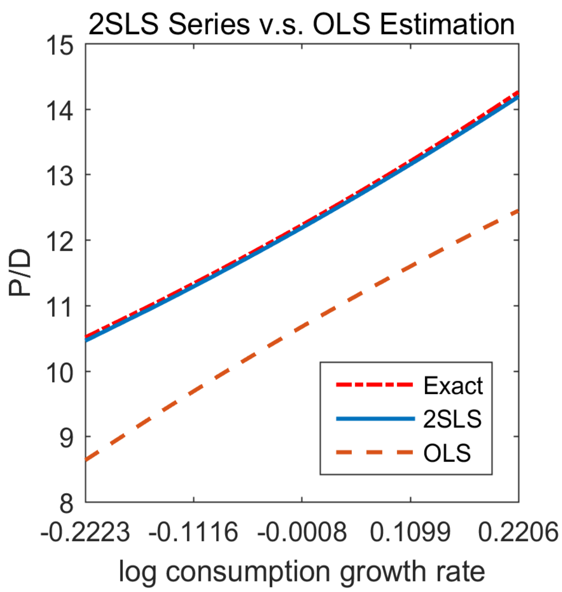
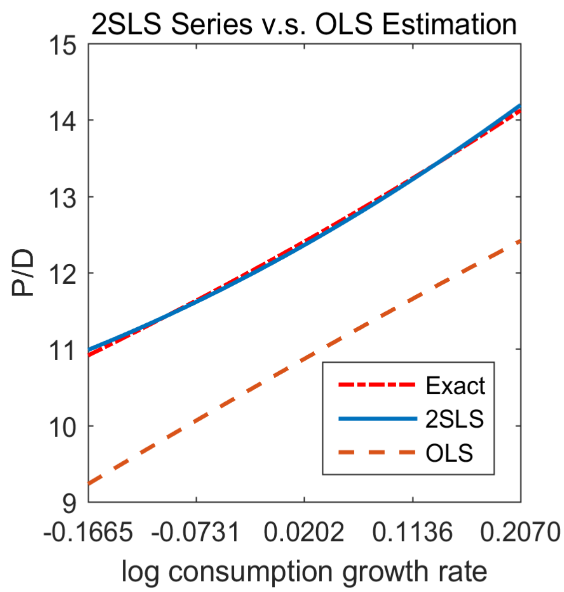


Figure 2.10: 2SLS Series vs OLS under DGP 5



Monte Carlo Simulations for Misspecified DGPs

In this section, we investigate possible consequences when the dynamics of state variables are misspecified using Mehra and Prescott's (1985) model. All current numerical solution methods require complete knowledge of the dynamics of state variables, whereas the true DGP of state variables in the real world is not completely known by empirical practitioners, possibly due to limited skill, time, or noisy observations. In practice, a proxy for the dynamics of state variables can be obtained via various techniques. For example, using simple rules of thumb, investors may obtain an estimated DGP which actually deviates from the true one in many dimensions (Cecchetti et al., 2000). Cecchetti et al. (2000) point out that this discrepancy between the true and subjective beliefs in the DGP of state variables is a key ingredient in addressing the equity premium puzzle. Even though it is common to encounter misspecified DGPs, very little attention has been paid to examining how asset pricing models can be affected when DGPs of state variables are misspecified. Therefore, we consider two empirically relevant true DGPs, and investigate how model implications can be altered when the price-dividend ratio function is solved based on a misspecified DGP via different solution methods.

DGP P.1: We have a true DGP, which follows an AR(1) process with a two-folded normally distributed disturbances:

$$X_{t+1} - \mu = \Gamma(X_t - \mu) + u^*_{t+1}, \text{ and}$$
$$f(u^*) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{u^{*2}}{2\sigma_1^2}}, \text{ if } u^* > 0, \\ \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{u^{*2}}{2\sigma_2^2}}, \text{ if } u^* \leq 0, \end{cases} \tag{2.32}$$

where $\sigma_1 = 3.48\%$, $\sigma_2 = 2\sigma_1$, $\beta = 0.96$, $\gamma = 1.5$, and $\Gamma = 0.8$. We assume that investors form a misspecified DGP as is described in Equation (2.33), which can match the true autocorrelation, and the first two true unconditional moments of X_t, while leaving all higher moments wrong.

$$\tilde{X}_{t+1} - \mu = \Gamma(\tilde{X}_t - \mu) + u_{t+1}, \text{ and } u_{t+1} \sim IIDN(0, \sigma_u^2), \tag{2.33}$$

where σ_u^2 is determined such that the variance of misspecified DGP is equal to the true variance, namely $\sigma_u^2 = var(X_t)(1 - \Gamma^2)$. Figure 11 shows how these two DGPs differ from each other using a time series plot.

DGP P.2: We have a threshold model for the true DGP, which is a nonlinear stationary process:

$$X_{t+1} = \begin{cases} \mu + \Gamma_1 X_t + u^*_{1,t+1},\ u^*_{1,t} \sim IIDN(0, \sigma_1^2) & \text{if } X_t > 0, \\ \mu + \Gamma_2 X_t + u^*_{2,t+1},\ u^*_{2,t} \sim IIDN(0, \sigma_2^2) & \text{if } X_t \leq 0 \end{cases} \tag{2.34}$$

where $\sigma_1 = 3.48\%$, $\sigma_2 = 2\sigma_1$, $\beta = 0.96$, $\gamma = 1.5$, $\Gamma_1 = 0.8$, and $\Gamma_2 = -0.139$. A misspecified DGP for such a process is as follows:

$$\tilde{X}_{t+1} - \mu = \bar{\Gamma}(\tilde{X}_t - \mu) + v_{t+1}, \text{ and } v_{t+1} \sim IIDN(0, \sigma_v^2), \tag{2.35}$$

where $\bar{\Gamma}$ and σ_v^2 are chosen so that it can match the autocorrelation with the true DGP. DGP P.2 explores a threshold structure, whose imporance has been widely acknowledged in many economic studies (e.g., Hong et al., 2012). In the true DGP, the state variable X_t is assumed to enjoy higher persistency level in mean and lower volatility when the consumption growth rate is positive, and will exhibit a mean-reverting pattern when the consumption growth rate is negative. Using a time series plot in Figure 13, we show that how the misspecified DGP fits the true data in practice.

We first solve the price-dividend ratio function using the perturbation and projection methods under the misperceived DGPs specified in Equation (2.33) and Example (2.35) respectively. We also estimate the price-dividend ratio function using the nonparametric 2SLS series regression approach, which does not require any specification of the DGP for state variables. Figures 12 and 14 compare the approximated price-dividend ratio functions from different solution methods under DGP P.1 and DGP P.2 respectively. In DGP P.1, it is clear that none of the discretization, perturbation or projection methods can provide a reasonable approximation for the true

Figure 2.11: DGP P.1

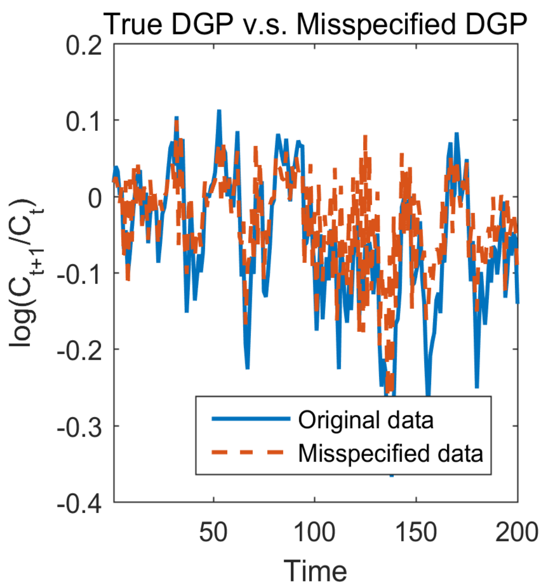
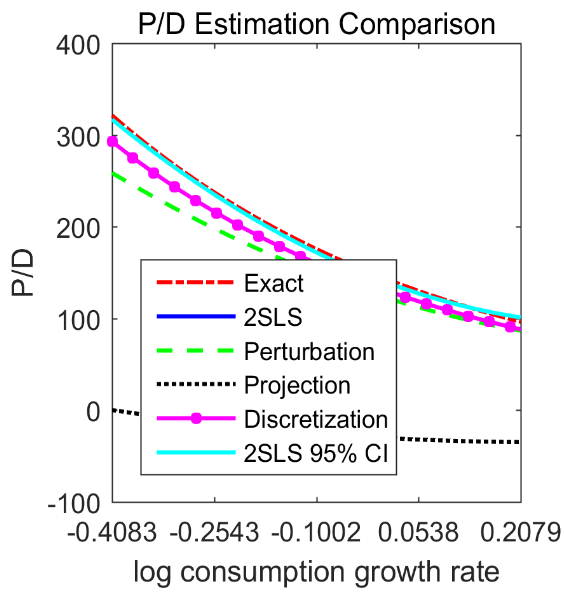


Figure 2.12: P/D under DGP P.1



price-dividend ratio function. Unlike other methods, the 2SLS series regression approach delivers an enormously superior estimation. As shown in Figure 14, the true DGP P.2 yields a convex price-dividend ratio function, which first decreases with a quadratic pattern and then increases almost linearly. When approximating such a price-dividend ratio function with a misspecified DGP, the linearization and projection methods both indicate a monotonically decreasing approximation, which falls far from the true values. However, the proposed 2SLS series regression approach can fit the true function almost perfectly over the entire domain because its solution is purely data-driven and does not involve any specification of the DGP of state variables.

Table 2.3 reports the first two moments of the risk-free and risky assets using different solution methods under DGP P.1 and DGP P.2. In a departure from the

Figure 2.13: DGP P.2

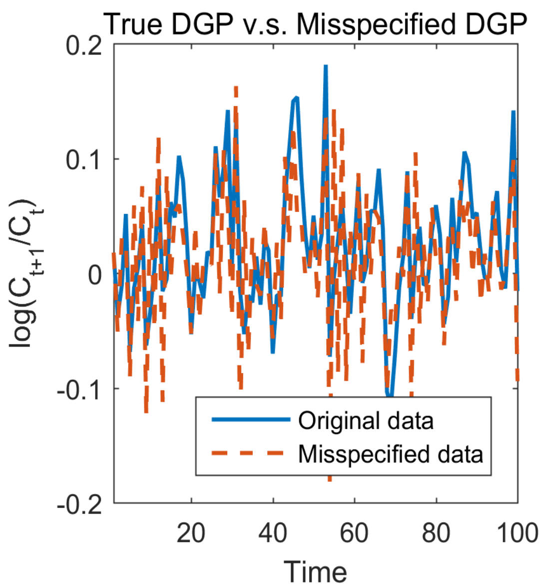


Figure 2.14: P/D under DGP P.2

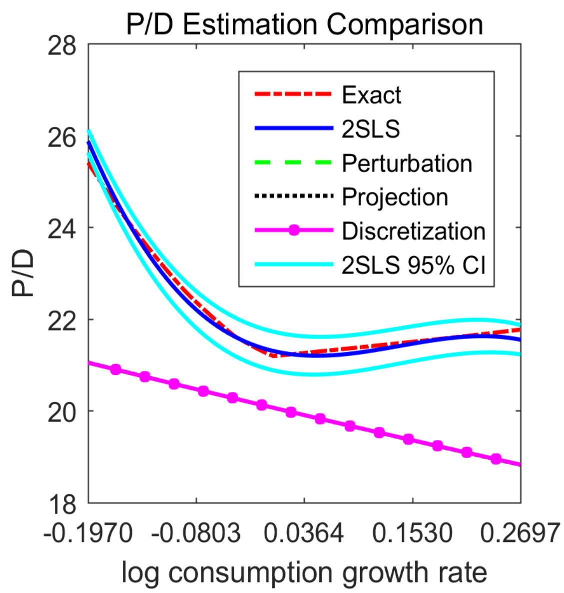


Figure 2.15: P/D under DGP 6

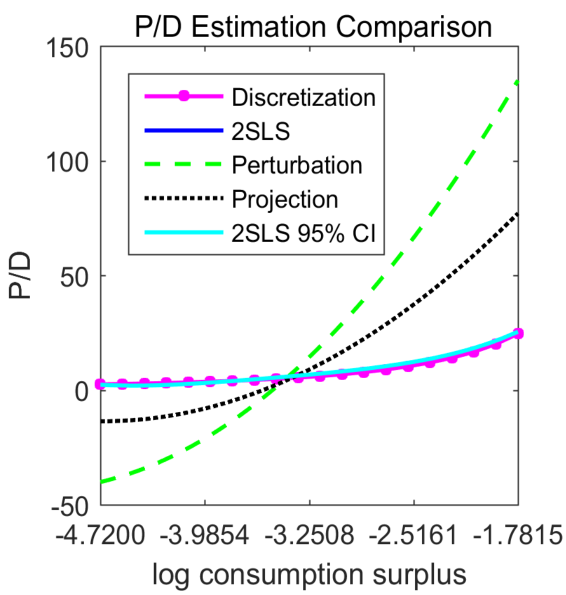
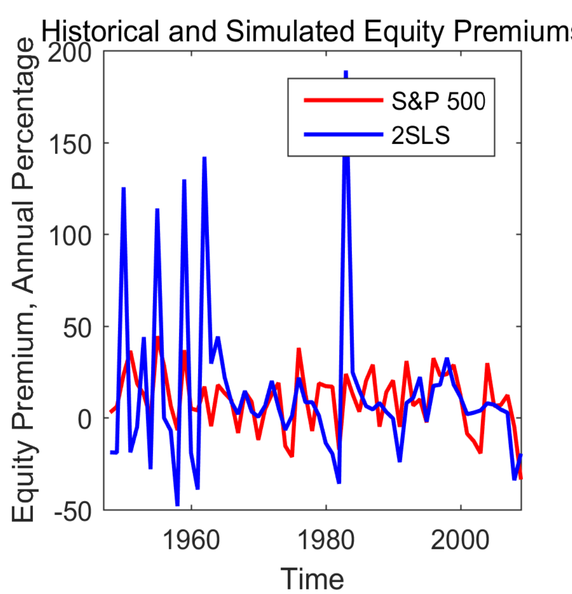


Figure 2.16: Historical and Simulated Equity Premiums under DGP 6



results shown in Table 2.2, all moments of the risk-free and risky assets are now affected by how the asset pricing model is solved in the presence of misspecified dynamics of state variables. In DGP P.1, because the discretization, perturbation and projection methods all suffer from the same type of misspecification error in the DGP of state variables, the mean and standard deviation of the risk-free return are the same, namely 25.808% and 3.809%, respectively. Further contaminated by their individual approximation errors, they generate different conclusions on the mean and standard deviation of the equity premium, all of which are wrong. In Contrast, the nonparametric 2SLS series regression approach can achieve an exact estimation of the first two moments of the risk-free return, namely -1.232% and 10.457%. The first two moments of the equity premium and the correlation between equities implied by the 2SLS series regression approach are also extremely close to the exact ones.

DGP P.2 considers a scenario with threshold nonlinear stationary state variables, which are wrongly modelled by a misspecified DGP. In this case, perturbation and projection methods all result in severely wrong conclusions about the mean and variance of the risk-free and risky assets and their correlation. In contrast, because the nonparametric 2SLS series regression approach does not rely on any specification of the dynamics of state variables, it exhibits superior accuracy in formulating equity moments in DGP P.2.

Using these two empirically relevant situations, we demonstrate that current numerical solution methods can further mislead model implications in the presence of misspecified dynamics of state variables. Therefore, when the DGP of state variables

Table 2.2: The First-two Moments of Asset Returns Using Different Solutions Methods for Price-dividend Ratios

	Model Solution Method	μ_f	σ_f	μ_{eq}	σ_{eq}	$\rho_{f,eq}$	MAD_f	MSE_f	MAD_{EL}	MSE_{EL}
		The first-two moments of assets					Solution Evaluation			
DGP 1	Analytic solution	9.683	1.334	0.352	4.465	0.0004	–	–	–	–
	Discretization (Tauchen)	9.683	1.334	0.352	4.465	0.0004	0.0000	0.0000	0.0000	0.0000
	2SLS series	9.683	1.334	0.350	4.469	0.0016	0.0034	0.0000	0.0006	0.0000
	Perturbation	9.683	1.334	0.352	4.464	0.0002	0.0001	0.0000	0.0001	0.0000
	Projection (Galerkin)	9.683	1.334	0.352	4.465	0.0002	0.0005	0.0000	0.0000	0.0000
DGP 2	Analytic solution	9.662	1.333	0.239	2.963	0.004	–	–	–	–
	Discretization (Tauchen)	9.662	1.333	0.239	2.963	0.004	0.0000	0.0000	0.0000	0.0000
	2SLS series	9.662	1.332	0.231	2.962	0.004	0.0033	0.0000	0.0017	0.0000
	Perturbation	9.662	1.333	0.239	2.965	0.003	0.0004	0.0000	0.0002	0.0000
	Projection (Galerkin)	9.662	1.333	0.239	2.963	0.004	0.0000	0.0000	0.0000	0.0000
DGP 3	Analytic solution	10.101	12.759	−1.110	14.508	0.002	–	–	–	–
	Discretization (Tauchen)	10.101	12.795	−1.109	14.508	0.002	0.0010	0.0000	0.0000	0.0000
	2SLS series	10.101	12.795	−1.130	14.662	−0.004	0.2148	0.0949	0.0814	0.0154
	Perturbation	10.101	12.795	−1.219	11.854	−0.164	1.5616	3.9342	0.1127	0.0425
	Projection (Galerkin)	10.101	12.795	−15.721	9.753	−0.891	33.1065	1128.7	1.9440	4.4946
DGP 4	Analytic Solution	10.378	12.884	0.603	6.393	0.014	–	–	–	–
	Discretization (Tauchen)	10.378	12.884	0.603	6.393	0.014	0.0000	0.0000	0.0000	0.0000
	2SLS series	10.378	12.884	0.626	6.389	0.015	0.0342	0.0012	0.0026	0.0000
	perturbation	10.378	12.884	0.601	6.389	0.012	0.0004	0.0000	0.0008	0.0000
	Projection (Galerkin)	10.378	12.884	2.756	6.980	0.201	2.4113	5.8160	0.1778	0.0384
DGP 5	Analytic solution	3.546	15.780	17.710	18.290	0.146	–	–	–	–
	Discretization (Tauchen)	3.546	15.780	17.710	18.290	0.146	0.0000	0.0000	0.0000	0.0000
	2SLS series	3.564	15.780	17.044	18.680	0.126	0.2688	0.0777	0.0325	0.0027
	Perturbation	3.546	15.780	22.508	14.195	−0.169	1.6606	2.9144	0.4390	0.2557
	Projection (Galerkin)	3.546	15.780	20.071	20.289	0.245	0.5922	0.3508	0.0676	0.0068

Notes: (i) All moments are in annual percentage. Moments from the analytic solution are based on the Burnside (1998)'s price-dividend ratios' algorithm. (ii) μ_f is the mean of risk-free asset, σ_f is the standard deviation of the risk-free asset, μ_{eq} is the mean of equity premium, σ_{eq} is the standard deviation of the equity premium, and $\rho_{f,eq}$ is the correlation between the risk-free and risky assets. (iii) DGPs are as specified in Table 3.1. Analytic benchmark results are based on the analytic solutions using Burside's (1998) method. (iv) All solution methods have order P chosen to 3. For the 2SLS series regression method, the price-dividend-ratio $\hat{f}_P$ has the projection coefficients estimated in Table 3.1.

is not fully specified, the newly proposed 2SLS series regression method will become a pivotal approach to obtain consistent estimate of the price-dividend ratio function, and construct the most reliable and accurate model implications.

Table 2.3: The First-two Moments of Asset Returns Using Different Solutions Methods for Price-dividend Ratios with Wrong DGP

		The first-two moments of assets				
	Model Solution Method	μ_f	σ_f	μ_{eq}	σ_{eq}	$\rho_{f,eq}$
DGP P.1	Correct solution	−1.232	10.457	−2.961	6.208	−0.587
	2SLS Series	−1.232	10.457	−2.583	5.168	−0.540
	Perturbation	7.808	11.530	−0.357	3.257	−0.069
	Projection (Garlerkin)	7.808	11.530	−5.321	11.690	−0.498
	Discretization (Tauchen)	7.808	11.530	−4.061	5.186	−0.086
DGP P.2	Correct solution	6.328	8.102	−0.263	8.042	−0.640
	2SLS Series	6.328	8.102	−0.250	7.959	−0.641
	Perturbation	7.740	3.678	0.547	4.953	0.023
	Projection (Garlerkin)	7.740	3.678	0.547	4.947	0.024
	Discretization (Tauchen)	7.740	3.678	0.547	4.953	0.023

Notes: (i) All moments are in annual percentage. Moments from the analytic solution are based on Burnside's (1998) price-dividend ratios algorithm. (ii) μ_f is the mean of risk-free asset, σ_f is the standard deviation of the risk-free asset, μ_{eq} is the mean of equity premium, σ_{eq} is the standard deviation of the equity premium, and $\rho_{f,eq}$ is the correlation between the risk-free and risky assets.

2.4.2 The Campbell and Cochrane (1999) Model

The Case with a Correctly Specified DGP

Campbell and Cochrane (1999) introduce levels of consumption habit into the classical CAPM by looking at the following maximization problem

$$\max_{\{C_t\}} E \sum_{t=0}^{\infty} \beta^t \frac{(C_t - H_t)^{1-\gamma} - 1}{1-\gamma}, \tag{2.36}$$

where H_t is the level of habit. The consumption habit S_t is assumed to be exogenous and determined by the history of aggregate consumption rather than the history of individual consumption. It follows that $S_t = (C_t - H_t)/C_t$. The log surplus consumption ratio $s_t \equiv lnS_t$ evolves as a heteroskedastic AR(1) process, namely

$$s_{t+1} = (1-\Gamma)\bar{s} + \Gamma s_t + l(s_t)(c_{t+1} - c_t - g), \tag{2.37}$$

where $\bar{s}$ is the steady state level and $l(s_t)$ is called the sensitivity function, specified as

$$l(s_t) = \begin{cases} \frac{1}{\bar{S}}\sqrt{1-2(s_t-\bar{s})} - 1, & \text{if } s_t \leq s_{max} = ln(\bar{S}) + \frac{1}{2}(1-\bar{S})^2, \\ 0, & \text{if } s_t \geq s_{max}. \end{cases} \tag{2.38}$$

The consumption growth $c_t = lnC_t$ is specified as an i.i.d. lognormal process $\Delta c_{t+1} = g + \nu_{t+1}$, where $\nu_{t+1} \sim i.i.d.N(0, \sigma^2)$. See Campbell and Cochrane (1999) for more detailed model description. The state variable is $X_t = s_t$. The price-dividend ratio f_t is embedded in the Euler equation:

$$f(X_t) = E\{\beta e^{-\gamma g} e^{-\gamma((\Gamma-1)(X_t-\bar{s}\bar{s})+(1-\lambda(X_t))\Delta c_{t+1})} e^{\Delta c_{t+1}}[1+f(X_{t+1})]|X_t\}. \tag{2.39}$$

Appealing accounts of the economic impact of consumption habits on the financial market were first reported by Campbell and Cochrane (1999) via the discretization method. At the same time, economists were concerned about functional form misspecification for most numerical solution methods. As a result, Campbell and Cochrane (1999) stimulate a series of conscientious studies exploring how the financial market will actually function if the model can be solved analytically. Chen et al. (2008) propose an analytic solution method for this model using the complex theory. However, this new analytic solution method only works with a small subset of the parameter space. Unfortunately, the exact model parameters used in Campbell and Cochrane's (1999) do not fall into the set that ensures an analytic solution. Therefore, Chen et al. (2008) evaluate the consumption habit model using some other parameter values as a proxy, and claim that the attractive equity premiums and Sharpe ratios cannot be matched if the model is solved analytically. Even though the discretization method is known to be challenged by interpolation errors and disappointing performance at extreme values, a fair judgement on model evaluation should be carried out at the exact parameter values used in Campbell and Cochrane (1999), instead of some close substitutes.

One appealing feature of our nonparametric 2SLS series regression approach is that it is not limited to parameter values of the DSGE model and can always obtain a consistent estimate of the unknown price-dividend ratio function corresponding to each set of model parameters. Therefore, in this section we conduct a model evaluation of the Campbell and Cochrane's (1999) model by comparing the 2SLS series regression approach to the discretization, perturbation and projection methods.

We first assume that empirical practitioners can correctly infer the dynamics of state variables, which implies that the DGP of state variables used when solving the model with different solution methods is correct. We estimate the price-dividend ratio f_t and predict equity premiums $E(R_{t+1} - R_{ft})$ using the 2SLS series regression. We use information criterion AIC to determine the order for p. Two simulation studies are carried out. Real data statistics are computed from two samples of the U.S. aggregate stock market. The long sample spans from 1890 to 2009, and the postwar sample includes data from 1949 to 2009. The equity data are Standard and Poor's 500 Price Index and Dividends. The risk-free rate is the return from the six-month commercial paper bought in January and rolled over in July. The empirical log consumption growth rates are computed from the real consumption per capita of nondurable goods and services. In DGP 6, we select values of parameters to match the postwar sample statistics. Parameters are set so that the annualized log consumption growth rate is 2.09% with a standard deviation equal to 1.81%, and the six-month Treasury bill has a 1.68% annualized return. In the second simulation study described by DGP 7, we match the long sample, where there is a 2.00% consumption growth rate with a 3.52% standard deviation, and a 1.91% return on the six-month Treasury bills. For comparison, we also numerically solve the model using the discretization, projection and perturbation methods.

Table 2.4 reports key statistics from the simulated data solved by different solution methods and the U.S. historical data respectively. In both simulation studies, data are designed to match the first two moments of the log consumption growth rates observed in the long and postwar U.S. data, and follow the true dynamics as described in DGP 6 and DGP 7. All data are simulated at annual frequency. In both DGP 6 and DGP 7, the autocorrelation of the state variable is $\Gamma = 0.87$, therefore we are conservative about the model implications from the projection method, whose accuracy is significantly discounted due to wide boundaries. The perturbation method suffers from functional-form misspecification errors, especially for tail areas. For DGP 6, Table 2.4 first reports significant differences in mean and variance of the equity premiums from different solution methods. The 2SLS series regression method indicates that the mean of equity premium is about 11.78% and its standard deviation is about 23.32%, which are both slightly smaller than that from the discretization method. However, the mean of equity premium from the perturbation

Table 2.4: Approximations for Consumption Claim Using Different Solution Methods

Panel A: DGP 6					
	Discretization	Projection	Perturbation	2SLS series	Postwar sample
$E(ln(C_{t+1}/C_t))$	2.09*	2.09*	2.09*	2.09*	2.09
$\sigma(ln(C_{t+1}/C_t))$	1.81*	1.81*	1.81*	1.81*	1.81
$E(R_f)$	1.68*	1.68*	1.68*	1.68*	1.68
$\sigma(R_f)$	0.38	0.38	0.38	0.38	2.92
$E(R_{t+1} - R_f)$	12.31	10.47	125.62	11.78	16.15
$\sigma(R_{t+1} - R_f)$	24.74	44.98	40.80	23.32	15.7
$E(P/D)$	11.03	36.57	51.97	11.43	32.61
$\sigma(P/D)$	3.29	16.07	24.18	3.51	16.69
$skewness(R_{t+1} - R_f)$	1.58	0.27	44.24	1.88	−0.22
$Kurtosis(R_{t+1} - R_f)$	6.37	233.11	1971.21	9.24	2.78
Panel B: DGP 7					
	Discretization	Projection	Perturbation	2SLS series	Long sample
$E(ln(C_{t+1}/C_t))$	2.00*	2.00*	2.00*	2.00*	2.00
$\sigma(ln(C_{t+1}/C_t))$	3.52*	3.52*	3.52*	3.52	3.52
$E(R_f)$	1.91*	1.91*	1.91*	1.91	1.91
$\sigma(R_f)$	0.19	0.19	0.19	0.19	5.78
$E(R_{t+1} - R_f)$	12.00	22.52	0.16	12.91	8.5
$\sigma(R_{t+1} - R_f)$	25.99	184.47	64.03	26.89	20.31
$E(P/D)$	11.59	23.94	22.47	11.45	26.64
$\sigma(P/D)$	2.49	4.89	5.14	2.54	13.81
$skewness(R_{t+1} - R_f)$	2.02	10.49	−5.58	2.32	−0.23
$Kurtosis(R_{t+1} - R_f)$	9.34	112.97	47.09	10.94	.32

Notes: (i) All returns are in annual percentage. (ii) R_{t+1} is the return of the risky asset. R_f is the return of the risk-free asset. The discretization column represents approximations from the value function iterated method. The projection method has order equal to three. The 2SLS method represents for the 2SLS series regression procedure. (iii) The long sample and postwar sample statistics are computed from the U.S. aggregate stock market. The long sample spans from 1890-2009 and the postwar sample spans from 1947-2009. The equity data are the Standard and Poor's 500 Price Index and Dividends. The risk-free rate is the return from six-month commercial paper bought in January and rolled over in July. The U.S. aggregate stock market data are downloaded from http://www.econ.yale.edu/ shiller/data.htm. (iv) Simulations are conducted in annual frequency. In the DGP 6 simulation study, parameters are given by: $g = 2.09\%$, $\sigma = 1.81\%$, $\Gamma = 0.87$, $\gamma = 2$ and β is chosen so that $E(R_f) = 1.68\%$. In the DGP 7 simulation study, parameters are given by: $g = 2.00\%$, $\sigma = 3.52\%$, $\Gamma = 0.87$, $\gamma = 2$ and β is chosen so that $E(R_f) = 1.91\%$. (v) $\star$ statistics are chosen to match the sample counterpart .

method is nearly 10 times larger than that from all other methods, and the variances of the equity premium from the projection and perturbation methods are both almost twice as large as that from the discretization and 2SLS series regression approaches. The skewness and kurtosis reported by the projection and perturbation methods are also very suspicious. When model parameters are set to match DGP 7, differences in the first two moments of the equity premiums from different solution methods are also dramatic. Overall, model implications from the 2SLS series regression method are close to that from the discretization method, and we believe that the 2SLS series regression method is the one that provides the most reliable conclusions because it is consistent and not affected by extreme values.

To further illustrate the possible reasons for this discrepancy, we plot the approximated price-dividend ratios in Figures 15 and 17. We expect positive price-dividend ratios. However, as shown in Figure 15, the approximated price-dividend ratio functions are significantly negative for some areas from the projection and perturbation methods. This is mainly due to approximation errors. On the other hand, approximated functions from the discretization and 2SLS series regression methods are very close. The main disagreement between these two solution methods is in the region with low surplus consumption ratios. As discussed by Chen et al. (2008), the solution from the discretization method is highly dependent on excessively large negative values for the dividend growth dynamics. This has a significantly adverse impact on the solution accuracy over areas with low surplus consumption ratios. Because our 2SLS series regression is robust to extreme values or tail distributions of state variables, it provides consistent estimates over the entire domain. In addition, the computational

Figure 2.17: P/D under DGP 7

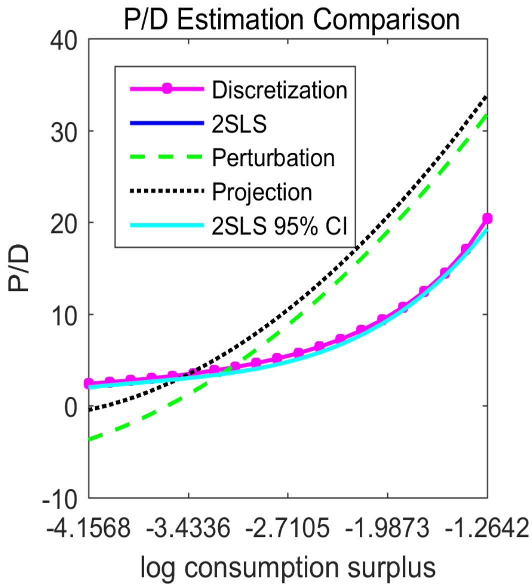
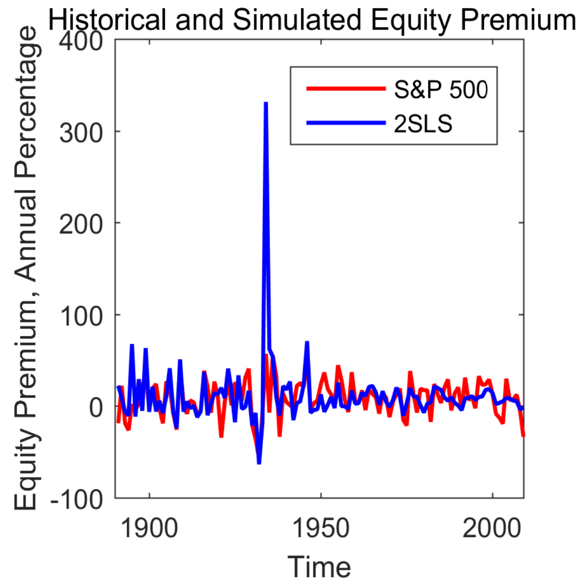


Figure 2.18: Historical and Simulated Equity Premiums under DGP 7



convenience of the 2SLS series method is not affected by model complexity. Similar results are reached in Figure 17. Specifically, both the projection and perturbation methods lead to disappointing approximations. However, the 2SLS series regression approach provides the most reliable estimation for the price-dividend ratio function over the entire domain.

Monte Carlo Simulations for Misspecified DGPs

Now, we explore situations when the dynamics of state variables are misspecified. Since the dynamics of the state variable, log consumption surplus, are not directly observable, we keep their structure as used by Campbell and Cochrane (1999) and only consider a situation where the log consumption growth rates are misspecified to some extent. Based on empirical observations, the log consumption growth rates have experienced various dynamics since they were first recorded. In the long sample, their volatility is about 1.81%, which is nearly half of that in the postwar period. Therefore, the true DGP that we construct in this section inherits this feature by assuming different volatilities in different regimes. We consider a true DGP of the state variable $\{s_{t+1}\}$ as follows:

DGP P.3: We have a true DGP, which follows a threshold AR(1) process with a two-folded normal distributed disturbances:

$$s^a_{t+1} = \begin{cases} (1-\Gamma)\bar{s}_1 + \Gamma s_t + l(s_t)(v_t - g_1),\ v_t \sim IIDN(g_1, \sigma_1^2) & \text{if } s_t > 0, \\ (1-\Gamma)\bar{s}_2 + \Gamma s_t + l(s_t)(v_t - g_2),\ v_t \sim IIDN(g_2, \sigma_2^2) & \text{if } s_t \leq 0, \end{cases} \tag{2.40}$$

where $v_t = c^a_{t+1} - c^a_t$, $g_1 = 2.09\%$, $\sigma_1 = 1.8\%$, $g_2 = 2\%$ and $\sigma_2 = 3.52\%$. $l(s_t)$, Γ, $\bar{s}_i$ for $i = 1, 2$ are the same as in Equations (2.37) and (2.38). In the meanwhile, we consider a misspecified DGP that fails to completely capture this dynamics as follows:

$$\tilde{s}_{t+1} = (1-\Gamma)\bar{s} + \Gamma\tilde{s}_t + l(\tilde{s}_t)(v_t - \frac{g_1+g_2}{2}),\ v_t \sim IIDN(\frac{g_1+g_2}{2}, \sigma_v^2), \tag{2.41}$$

where $v_t = c_{t+1} - c_t$, $\sigma_v^2 = var(s_t)(1-\Gamma^2)$, $l(\tilde{s}_t)$, Γ and $\bar{s}$ are the same as in Equations (2.37) and (2.38). As can be seen, the misspecified DGP shares the same autocorrelation and first two moments of the true DGP. However, we will show that this mild mistake in the specification of the state variable will have a significantly adverse impact on model conclusions. The estimated price-dividend ratio function from the 2SLS series regression approach is purely data-driven and does not require

Table 2.5: The First-two Moments of Asset Returns Using Different Solutions Methods for Price-dividend Ratios with Wrong DGP

		The first-two moments of assets				
	Model Solution Method	μ_f	σ_f	μ_{eq}	σ_{eq}	$\rho_{f,eq}$
DGP P.3	Correct Solution	1.577	0.698	22.095	72.912	−0.075
	2SLS Series	1.577	0.698	18.735	93.681	−0.081
	Perturbation	1.642	0.303	9.149	103.997	−0.053
	Projection (Garlerkin)	1.642	0.303	14.000	221.636	−0.063
	Discretization (Tauchen)	1.642	0.303	14.480	26.411	−0.220

Notes: (i) All moments are in annual percentage. Moments from the analytic solution are based on Burnside's (1998) price-dividend ratios algorithm. (ii) μ_f is the mean of risk-free asset, σ_f is the standard deviation of the risk-free asset, μ_{eq} is the mean of equity premium, σ_{eq} is the standard deviation of the equity premium, and $\rho_{f,eq}$ is the correlation between the risk-free and risky assets.

any knowledge of the dynamics of state variables. Therefore, all model implications from this newly proposed method are immune to model misspecification of the state variables.

Table 2.5 reports the first two moments of the risk-free and risky assets using the 2SLS series regression, discretization, projection and perturbation methods in the presence of misspecified DGP of state variables. The mean and standard deviation of the risk-free asset reported by all three numerical solution methods are 2.823% and 8.886%. However, the correct values, as obtained from the 2SLS series regression method, are 1.614% and 0.458%. Meanwhile, the first two moments of equity premiums obtained from the three numerical solution methods are not only significantly different, they also contradict each other, falling far afield of the correct values reported by our 2SLS series regression method.

Figure 2.19: DGP P.3

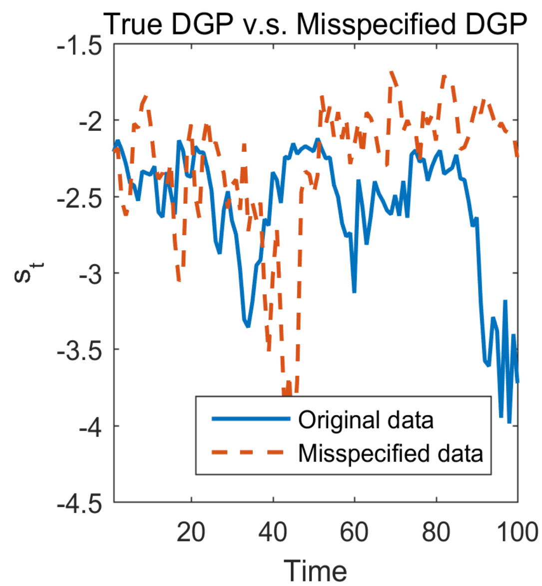
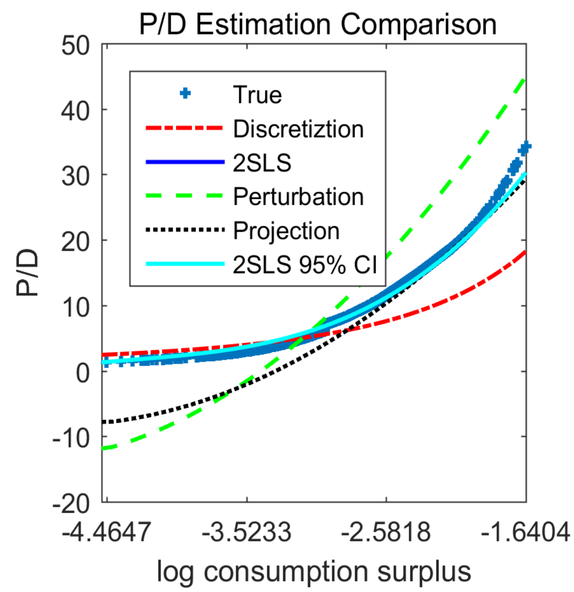


Figure 2.20: P/D under DGP P.3



We plot the price-dividend ratio function from different solution methods in Figure 19. The estimated function from the 2SLS series regression method is first slightly concave and then eventually becomes convex. Approximations from the other three numerical solution methods all fail to satisfactorily capture this underlying true process. This poor approximation of the price-dividend ratio function directly leads to misleading results related to equity premiums.

2.5 Conclusion

The explanatory capability of the CAPM heavily relies on the solution accuracy of price-dividend ratios as a function of state variables. Poor approximations due to functional form misspecification errors of the price-dividend ratio function and misspecified DGP of state variables may seriously discredit economic interpretations. Built upon a classical CAPM framework, this paper has demonstrated the extent to which a price-dividend ratio solution that suffers from functional form misspecification and misspecified DGP of state variables can adversely affect the explanation for equity premiums. While existing popular numerical solution approaches are appealing due to their flexibility and wide application, they all suffer from functional form misspecification, require a specification for the DGP of state variables, and involve tedious computation, especially when the CAPM is complex. In this paper, we have proposed a nonparametric 2SLS series regression procedure to estimate and evaluate price-dividend ratios as a function of state variables in the CAPM with time-separable utility functions. The unknown price-dividend ratio function is specified recursively over different time periods under rational expectations, and state variables are assumed to follow stationary Markov processes. Since the recursive nature of the price-dividend ratio function induces an endogeneity bias, we propose an 2SLS series regression method, which has a convenient data-based closed-form solution regardless of model complexity, thus making the implementation particularly easy in practice. Most importantly, our method is free of endogeneity biases and functional form misspecification when the sample size goes to infinity. In addition, our 2SLS series estimator is shown to be consistent and asymptotically normal.

When solving unknown functions such as price-dividend ratios, all existing numerical solution methods in the literature impose a pre-specified functional form, with all involved parameters calibrated via either matching coefficients or numerical integration techniques. Except for functional form misspecification, price-dividend ratios derived via these methods are further limited in two important dimensions. First, all existing numerical solution methods for price-dividend ratios are obtained by assuming a specification for the DGP of state variables, which therefore may suffer from model misspecification. Second, because no estimation is involved in solving for price-dividend ratios, they do not have any statistical properties. As a result, practitioners are unable to obtain a rigorous inference on model evaluation. In contrast, one of the most important features of the 2SLS series regression is that the solution of price-dividend ratios does not require any prior knowledge of the data generating process of state variables. It works with all state variables that belong to a broad class of stationary Markov processes. The price-dividend ratios estimated by our procedure not only have a closed-form solution that does not depend on model complexity, but also have nice statistical properties which facilitate rigorous inference.

Our empirically relevant simulation studies show that the 2SLS series regression method performs reasonably well in finite samples and under various parametrizations. It outperforms existing popular solution methods such as the permutation, projection and discretization methods for nonlinear price-dividend ratio functions and for tail distributions of state variables when the DGP of state variables is known. Due to limited time and skills, the perceived DGP of state variables may differ dramatically from its true underlying processes. This paper investigates several empirically relevant setups, and we find that all current numerical solution methods result in even worse approximations and contradictory model implications in the presence of misspecified DGP of state variables. In contrast, the 2SLS series regression approach does not require any knowledge on the DGP of state variables and works with a broad class of stationary Markov processes. Therefore, it will become a pivotal tool to construct reliable and correct conclusions on model implications and evaluation for DSGE models.

Our approach can be generalized in several directions. Meghir and Pistaferri (2004) model the conditional variance of the income shocks as a parsimonious ARCH process. It helps them achieve significant improvement in understanding household counterfactual consumptions by capturing education- and time-specific differences in the stochastic process for earnings and for measurement error. By applying this newly proposed method, we can learn the extent to which income risks affect equity prices without modelling the stochastic process of income risks. Also, by incorporating empirical observations of state variables and avoiding model misspecification, we can better understand how monetary and fiscal policies will actually function in the real economy. In addition, this method can be extended to DSGE models in the production economy, where the log linearization method is widely used (Zietz, 2006). Lastly, under a system of multiple Euler equations, we must solve multiple unknown functions rather than just the price-dividend ratio function (Epstein and Zin, 1989). It is important to solve all unknown functions accurately because possible functional form misspecification from one solution may be amplified and adversely affect the others, eventually seriously discrediting model implications. Our 2SLS series regression approach can be extended to this more general and complex setup, eliminating all possible functional form misspecification in large samples. This newly proposed functional estimation method will facilitate a more reliable understanding of existing DSGE models that are now widely used in both macroeconomics and finance.

CHAPTER 3

A NONPARAMETRIC GMM SERIES APPROACH TO SOLVING MULTI-EQUATION ASSET PRICING MODELS WITH RECURSIVE PREFERENCES

3.1 Framework

We introduce a nonparametric GMM series estimation procedure in the context of Epstein and Zin's (1989) model, which is recognized as one of the most influential papers that provide important insights in the literature. There is an infinitely-lived representative agent who maximizes the expected life-time utility V_t at time t, namely

$$
\begin{aligned}
& V_t = \{(1-\beta)C_t^{\frac{1-\gamma}{\theta}} + \beta[E_t(V_{t+1}^{1-\gamma})]^{\frac{1}{\theta}}\}^{\frac{\theta}{1-\gamma}} \\
\text{s.t.} \quad & C_t + P_{t+1}\theta_{t+1} + Q_t b_{t+1} = b_t + (D_t + P_t)\theta_t,
\end{aligned}
\tag{3.1}
$$

where C_t is consumption level at time t, β is the constant time discount factor, γ is the coefficient of relative risk aversion level, η is the intertemporal elasticity of substitution, and $\theta = \frac{1-\gamma}{1-\frac{1}{\eta}}$. When $\eta = 1$, we have the Mehra and Prescott's (1985) CAPM as a special case of Epstein and Zin (1989). Campbell and Cochrane (2000) point out that most asset pricing models can be derived as various specifications of this framework. Therefore, we use this model as the basis to introduce our nonparametric GMM series method. By taking derivatives, we obtain a system of Euler equations as follows:

$$\begin{cases} E[\beta^{\theta}(\frac{C_{t+1}}{C_t})^{-\frac{\theta}{\eta}} R_{w,t+1}^{\theta-1} R_{i,t+1}|I_t] = 1, \\ E_t[\beta^{\theta}(\frac{C_{t+1}}{C_t})^{-\frac{\theta}{\eta}} R_{w,t+1}^{\theta}|I_t] = 1, \end{cases} \tag{3.2}$$

where I_t denotes all the information available at time t, and $R_{i,t+1}$ and $R_{w,t+1}$ are the returns of the risky asset and aggregate wealth at time $t+1$. Specifically,

$$\begin{cases} R_{w,t+1} = \frac{W_{t+1}}{W_t - C_t} = \frac{\frac{W_{t+1}}{C_{t+1}}}{\frac{W_t}{C_t} - 1} \frac{C_{t+1}}{C_t}, \\ R_{i,t+1} = \frac{P_{t+1}+D_{t+1}}{P_t} = \frac{\frac{P_{t+1}}{D_{t+1}}+1}{\frac{P_t}{D_t}} \frac{D_{t+1}}{D_t}. \end{cases} \tag{3.3}$$

Without abuse of notations, we let $h_t = \{f_t, g_t\}$, where $f_t \equiv ln(\frac{P_t}{D_t})$ and $g_t \equiv ln(\frac{W_t}{C_t})$ be the logarithm of the price-dividend ratio and wealth-consumption ratio functions. It is important to solve these two recursively specified unknown functions accurately so as to ensure reliable conclusions on asset returns. Expressing asset returns using h_t, we can rewrite Euler equations (3.2) as follows:

$$\begin{cases} E[\beta^{\theta}(\frac{C_{t+1}}{C_t})^{\theta-1-\frac{\theta}{\eta}}(\frac{D_{t+1}}{D_t})(\frac{e^{g_{t+1}}}{e^{g_t}-1})^{\theta-1}(e^{f_{t+1}}+1) - e^{f_t}|I_t] = 0, \\ E[\beta^{\theta}(\frac{C_{t+1}}{C_t})^{\theta-\frac{\theta}{\eta}}(\frac{e^{g_{t+1}}}{e^{g_t}-1})^{\theta} - 1|I_t] = 0. \end{cases} \tag{3.4}$$

Let $\{X_t\}$ summarizes the law of motions of all state variables. In the present setup, $X_t = (\frac{C_t}{C_{t-1}}, \frac{D_t}{D_{t-1}})'$ and $I_t = \{X_t, X_{t-1}, \cdots, X_0\}$. Given the dynamics of state variables, the solutions of the price-dividend ratio and wealth-consumption ratio functions of state variables, which are recursively specified in Equation (3.4), are of our central interest.

Given rapid development and increasing model complexity in macroeconomics and finance, the desired solution method is expected to work with as many empirically relevant setups as possible. Cecchetti and Lam consider a Markov-regime switching process for state variables. Campbell and Cochrane (1999) specify a nonlinear autoregressive process state variable which can incorporate consumption habits. Tallarini (2000) consider a situation where state variables are hidden among some noise-driven observations. Bansal and Yaron (2004) and Bansal, Kiku and Yaron (2012) further investigate the role played by potential long-run risks in addressing equity premiums using an autoregressive conditional heteroskedasticity (ARCH) model for state variables. Because investors are prone to extrapolate historical data when predicting further stock performance, serially dependent state variables in DSGE models are strongly suggested in the attention-drawing survey conducted by Greenwood and Shleifer (2014). Furthermore, Hansen and Scheinkman (2012) argue the generality of Markov processes of state variables in DSGE models. Therefore, built upon a broad class of stationary, non-Gaussian and Markov multi-dimensional state variables $\{X_t\}$, this paper introduces a nonparametric GMM series procedure to solve the price-dividend ratio function and wealth-consumption ratio function simultaneously in the Epstein and Zin's (1989) model. These two unknown functions are recursively specified under the rational expectation in a system of Euler equations, and will be estimated simultaneously. While capturing the relationship among different unknown functions using the variance-covariance matrix and the optimal weighting matrix, the nonparametric GMM series method can significantly improve solution efficiency. Moreover, this new method use a nonparametric series model for each unknown function of state variables, and so is free of functional form misspecification when the sample size $T \to \infty$, no matter how complex the DSGE model is. Unlike all existing numerical solution methods, our method does not involve any specification and estimation of the unknown dynamics of state variables. It is general and flexible enough to facilitate analysis of a wide variety of DSGE models, because it can be easily implemented without alternations.

3.2 Nonparametric GMM Series Estimation

Let F_t be the cumulative distribution function of state variables X_t. Let $\{X_{t+1}\}$ be a vector of Markov processes that the conditional probability function of X_t given the information set $I_t \equiv \{X_t, X_{t-1}, \cdots\}$ only depends on its previous lagged variable X_t.

Assumption 3.2.1. *The state variables X_t follows a Markov process with a positive density dF_t/dX_t that is continuous almost everywhere on $\mathbb{X}$.*

Therefore, under the Markov assumption, the original simultaneous Euler equations (3.4), which recursively specify unknown functions $[f^o(x), g^o(x)]$, can be represented as follows:

$$\begin{cases} E\big\{\big[\beta^\theta(\frac{C_{t+1}}{C_t})^{\theta-1-\frac{\theta}{\eta}}(\frac{D_{t+1}}{D_t})[\frac{e^{g(X_{t+1})}}{e^{g(X_t)}-1}]^{\theta-1}[e^{f(X_{t+1})}+1]-e^{f(X_t)}\big]|X_t\big\}=0, \\ E\big\{\big[\beta^\theta(\frac{C_{t+1}}{C_t})^{\theta-\frac{\theta}{\eta}}[\frac{e^{g(X_{t+1})}}{e^{g(X_t)}-1}]^{\theta}-1\big]|X_t\big\}=0. \end{cases} \tag{3.5}$$

Consider $h^o \equiv \{f^o, g^o\}$ is the pair of functions that uniquely solves Equation (3.5). As suggested by Cui and Hong (2016), who propose a nonparametric 2SLS series procedure for a single Euler equation, we shall estimate the set of recursively specified unknown functions $\{f^o, g^o\}$ using a global series method instead of local approximation techniques. Local constant and local polynomial approximations are among the most popular local estimation methods in nonparametric analysis (e.g., Fan and Gijbels, 1996). These methods are based on local Taylor expansions at some steady points. Because different unknown functions may possess various properties, it is difficult to determine an appropriate common point for local Taylor expansions for all unknown functions. In addition, embeded in Equations (3.4), $f_t = f(X_t)$, $g_t = g(X_t)$, $f_{t+1} = f(X_{t+1})$ and $g_{t+1} = g(X_{t+1})$ are all unknown and must be estimated simultaneously at each time t. However, the actual distance $X_{t+1} - X_t$ varies with time and can be substantially large. Therefore, it is challenging to pin down a suitable point in all time periods for local Taylor expansions as well using local estimation methods.

Without seeking local centers for each unknown function in each time period t, we consider a global series approximation in the present context. Let $\{\phi_j\}_{j=1}^{\infty}$ and $\{\psi_j\}_{j=1}^{\infty}$ be two sequence of complete basis functions. To estimate unknown functions $\{f^o, g^o\}$ in Equations (3.5), we use series approximations

$$\begin{cases} f_p(x) = \sum_{j=1}^{p} a_j \phi_j(x) = \phi^p(x)' a^p, \\ g_q(x) = \sum_{j=1}^{p} b_j \psi_j(x) = \psi^q(x)' b_q, \end{cases} \tag{3.6}$$

where $a^p = (a_1, \cdots, a_p)'$ and $b^q = (b_1, \cdots, b_q)'$. The orders p and q must grow to infinity as the sample size $T \to \infty$. Now we impose the following mild conditions on p, q and basis functions.

Assumption 3.2.2. *Let $\{\phi\}_{j=1}^{\infty}$ and $\{\psi\}_{j=1}^{\infty}$ be two complete basis functions defined on a normed space $\mathbb{X}$. Suppose $f \in L^2$ and $g \in L^2$ are measurable and continuously differentiable up to order $d \geq 0$. Let truncated series $f_p(x) = \sum_{j=1}^{p} a_j\phi_j(x)$ and $g_q = \sum_{j=1}^{p} b_j\psi_j(x)$, where $p \equiv p(T) \to \infty$ and $q \equiv q(T) \to \infty$ as the sample size $T \to \infty$ such that (i) for an integer $d \geq 0$, there are $s > 0$, $a^p = (a_1, \cdots, a_p)'$ and $b^q = (b_1, \cdots, b_q)'$ so that $|f - f_p|_d = \mathbb{O}(p^{-s})$ and $|g - g_q|_d = \mathbb{O}(q^{-s})$; (ii) Letting $k = \max(p, q)$, $\sqrt{T}k^{-s} \to 0$.*

Note that $\{\phi\}_{j=1}^{\infty}$ and $\{\psi\}_{j=1}^{\infty}$ may be different. Assumption 3.2.2 (i) is a rate condition at which the approximation biases $f^o(x) - f_p(x) = \sum_{j=p+1}^{\infty} a_j\varphi_j(x)$ and $g^o(x) - g_q(x) = \sum_{j=q+1}^{\infty} b_j\phi_j(x)$ vanish to zero as $p, q \to \infty$ as the sample size $T \to \infty$. To control the approximation biases of the series estimators $\hat{h}(x) \equiv [\hat{f}_p(x), \hat{g}_q(x)]$, Assumption 3.2.2 (ii) further requires that both $p = p(T)$ and $q = q(T)$ go to infinity at a rate slower than $\sqrt{T}$ but faster than $T^{-\frac{1}{2s}}$.

Given various properties that different unknown functions may possess, we need to choose basis functions appropriately. We first enumerate some scenarios when $\mathbb{X}$ is compact. Suppose f^o and g^o are periodic functions and continuously differentiable of order $d \in \mathbb{N}$ over a compact support, say $\mathbb{Q} \equiv [a, b]^n$, where a and b are finite constants with $a < b$, and n is the dimension of state variables X_t. Then we can consider the trigonometric series on $\mathbb{Q}$. Specifically, we consider the following Fourier series approximations:

$$\begin{cases} f_p(x) = d_0^f + \sum_{i=1}^{I_n}\sum_{j=1}^{J_n}\{d_{ij}^f cos(jk_i'x) + w_{ij}^f sin(jk_i'x)\} = \sum_{j=1}^{p} a_j\phi_j(x), \\ g_q(x) = d_0^g + \sum_{i=1}^{I_n}\sum_{j=1}^{J_n}\{d_{ij}^g cos(jk_i'x) + w_{ij}^g sin(jk_i'x)\} = \sum_{j=1}^{q} b_j\psi_j(x), \end{cases} \tag{3.7}$$

where $I_n,\ J_n \in \mathbb{N}$, d_0^f, d_{ij}^f, d_0^g, d_{ij}^g, w_{ij}^f and $w_{ij}^g \in \mathbb{R}$. $k_i \in \mathbb{K}_T \equiv \{k_i : i = 1, \cdots, I_n\}$ is an elementary multi-index, a $n \times 1$ vector of integers. For the construction of k_i, see Gallant (1981). It is easy to show that Assumption 3.2.2 will be automatically satisfied when setting $s = d/n$.

The periodicity assumption made on $h^o = \{f^o, g^o\}$ appears strong. Relaxing it will result in boundary effects, which become an important issue to resolve in series approximations(Gallant and Souza, 1991). Gallant and Souza (1991) introduce a Flexible Fourier Form (FFF) series to improve the performance of boundary regions. Hong and White (1995) further apply it in a nonparametric testing framework with insightful conclusions on its asymptotic results. Given the appealing advantages of the FFF series in the boundary regions, we employ it as an example to work with non-periodic functions. We consider a FFF series on a compact support $\mathbb{Q} = [\nu, 2\pi - \nu]^n$ with any small $\nu > 0$:

$$\begin{cases} f_p(x) = d_0^f + \sum_{i=1}^{l} b_r^f x_i + \sum_{i=1}^{l}\sum_{j=1}^{i} c_{ij}^f x_i x_j + \sum_{i=1}^{I_n}\sum_{j=1}^{J_n}\{d_{ij}^f cos(jk_i'x) + w_{ij}^f sin(jk_i'x)\} \\ \qquad = \sum_{j=1}^{p} a_j \phi_j(x), \\ g_q(x) = d_0^g + \sum_{i=1}^{l} b_r^g x_i + \sum_{i=1}^{l}\sum_{j=1}^{i} c_{ij}^g x_i x_j + \sum_{i=1}^{I_n}\sum_{j=1}^{J_n}\{d_{ij}^g cos(jk_i'x) + w_{ij}^f sin(jk_i'x)\} \\ \qquad = \sum_{j=1}^{q} b_j \psi_j(x), \end{cases} \tag{3.8}$$

where $(a_1, \cdots, a_p) = (d_0^f, a_{(0)}, a_{(1)}, \cdots, a_{(I_n)})$, $a_{(0)} = (b_1^f, \cdots, b_d^f, c_{11}^f, c_{12}^f, \cdots, c_{dd}^f)$, $a_{(i)} = (d_{i1}^f, w_{i1}^f, \cdots, d_{iJ_n}^f, w_{iJ_n}^f)$, $(b_1, \cdots, b_p) = (d_0^g, b_{(0)}, b_{(1)}, \cdots, b_{(I_n)})$, $b_{(0)} =$

$(b_1^g, \cdots, b_d^g, c_{11}^g, c_{12}^g, \cdots, c_{dd}^g)$, and $b_{(i)} = (d_{i1}^g, w_{i1}^g, \cdots, d_{iJ_n}^g, w_{iJ_n}^g)$. For more discussion of the FFF series, see Gallant and Souza (1991).

Next, we consider series approximations using regression splines. Newey (1994) suggests using B-splines because it helps alleviate collinearity problems significantly. Without loss of generality, let $\mathbb{Q} = [0,1]^n$, $\Delta = \{s_i\}_{i=1}^k$ with $0 = s_1 < s_2 \cdots < s_{k+1} = 1$ be a partition of $\mathbb{Q}$ into k intervals $I_j = [s_j, s_{j+1})$ where $j \in \{1, \cdots, k-1\}$, and $I_k = [s_k, s_{k+1}]$. The space of polynomial splines of order $w \in N$ with knots $s_1, \cdots, s_k$ is defined as,

$$\begin{cases} f_p : \mathbb{X}_s \to \mathbb{R}, f_p(x) = \sum_{j=1}^p a_j \phi_j(x) \quad \text{for } x \in I_j, \\ \text{where } f_p(x) \text{ is a } w\text{-th order polynomial} \in C^{w-2} \text{ at } s_j,\ j = 1, \cdots, k. \quad \phi_j : \mathbb{Q} \to \mathbb{R}, a_j \in \mathbb{R}, \\ g_q : \mathbb{X}_s \to \mathbb{R}, g_q(x) = \sum_{j=1}^q b_j \phi_j(x) \quad \text{for } x \in I_j, \\ \text{where } g_q(x) \text{ is a } w\text{-th order polynomial} \in C^{w-2} \text{ at } s_j,\ j = 1, \cdots, k. \quad \varphi_j : \mathbb{Q} \to \mathbb{R}, b_j \in \mathbb{R}. \end{cases} \tag{3.9}$$

A direct choice for $\{\phi_j\}_{j=1}^{\infty}$ and $\{\varphi_j\}_{j=1}^{\infty}$ is the normalized wth order B-splines $\{N_i^w\}$ with knots $s_j, \cdots, s_{j+w}$ that satisfy $\sum_{i=j+1-w}^{j} N_i^w(x) = 1$ for all $s_j \leq x < s_{j+1}$.

Last, we discuss choices of basis functions when $h^o = \{f^o, g^o\}$ are defined over an open subset $\mathbb{Q} \equiv (a,b)^n$. A considerable number of DSGE models in macroeconomics and finance actually belong to this scenario. For example, when $\mathbb{Q} \equiv (0, \infty)^n$, it covers DSGE models with state variables following Gamma and F-distributions. In addition, when $\mathbb{Q} \equiv (-\infty, \infty)^n$, the nonparametric GMM series approach will incorporate state variables with normal and t-distributions. To facilitate our analysis in these scenarios, we consider convergence in the weighted Sobolev norm $||\cdot||_{d,w}$ in the weighted Sobolev space H_w^d, where

$$H_w^d = \{f : \mathbb{R} \to \mathbb{R} | f, f^{'}, \cdots, f^{(d)} \in L_w^2\}, \quad d \in \mathbb{N}, \tag{3.10}$$

and

$$L_w^2 = \{f : \mathbb{R} \to \mathbb{R} | \int_a^b f(x)^2 w dx < \infty\}. \tag{3.11}$$

The weighted Sobolev norm $|| \cdot ||_{d,w}$ in H_w^d is given by

$$||f||_{d,w} = \sum_{j=0}^{d} ||f^{(j)}||_{0,w}, \quad \text{with} \quad ||f||_{0,w} = [\int_{-\infty}^{\infty} |f(x)|^2 w(x) dx]^{\frac{1}{2}}. \tag{3.12}$$

When setting $w(x) = 1$ and $d = 0$, we have $||f||_2 = [\int_{-\infty}^{\infty} |f(x)|^2 dx]^{\frac{1}{2}}$ as a special case of the weighted Sobolev norm.

Choices of basis functions depend on the type of the domain $\mathbb{Q}$ and weighting function w. We first consider the situation of $\mathbb{Q} \equiv (0, +\infty)^n$, where it becomes to $(-\infty, 0)^n$ when we put state variables to $\{-X_t\}$. We consider generalized Laguerre series with weighting function $w_\alpha(x) = x^\alpha e^{-x}$ for some $\alpha > -1$. The generalized Laguerre polynomials of order p and q are defined as

$$\begin{cases} f_p(x) = \sum_{j=1}^{p} a_j \phi_j(x), \quad \text{where } \phi_j(x) = \frac{1}{j!} x^{-\alpha} e^x \frac{\partial^j x^{j+\alpha} e^{-x}}{\partial x^j}, \ j = 0, 1, \cdots, p, \\ g_q(x) = \sum_{j=1}^{p} b_j \psi_j(x), \quad \text{where } \psi_j(x) = \frac{1}{j!} x^{-\alpha} e^x \frac{\partial^j x^{j+\alpha} e^{-x}}{\partial x^j}, \ j = 0, 1, \cdots, q. \end{cases} \tag{3.13}$$

Guo et al. (1991) generalizes and improves the result on Laguerre approximations. For any $0 \leq s \leq d$, $||f_p - f||_{s,w(\alpha)} = \mathbb{O}(p^{\frac{s-d}{2n}})$ and $||g_q - q||_{s,w(\alpha)} = \mathbb{O}(q^{\frac{s-d}{2n}})$.

Then, we consider the situation of $\mathbb{Q} \equiv (-\infty, +\infty)^n$. There is a rich literature in macroeconomics and finance that incorporates stochastic processes on the entire real space. The Hermite series is considered as an appealing tool. Gallant (1981) use the Hermite series for maximum likelihood estimation. The Hermite series is further employed by Gallant and Souza (1991) to estimate the conditional density function of Euler equations in CAPMs with separate utilities. Aït-Sahalia (2002) considers the Hermite series in estimating discretely sampled diffusion processes. Ait-Sahalia et al. (2012) test jump diffusion models nonparametrically via the Hermite series. Given its popularity and wide application in estimating various functions, we consider Hermite series approximations as

$$\begin{cases} f_p = \sum_{j=1}^{p} a_j \phi_j(x), \text{where } \psi_j(x) = e^{-x^2} H_j(x),\ H_j(x) = (-1)^j e^{t^2 \frac{d^j e^{-t^2}}{dt^j}}, \text{ and } w(x) = e^{-x^2}, \\ g_q = \sum_{j=1}^{q} b_j \psi_j(x), \text{where } \phi_j(x) = e^{-x^2} H_j(x),\ H_j(x) = (-1)^j e^{t^2 \frac{d^j e^{-t^2}}{dt^j}}, \text{ and } w(x) = e^{-x^2}. \end{cases} \tag{3.14}$$

Note that two basis functions need not to be the same.

After making an appropriate decision on choices of basis functions, we obtain a pair of truncated series approximations $h_l \equiv \{f_p, g_q\}$ for $h^o \equiv \{f^o, g^o\}$, where $l = p + q$. Using the orthogonality conditions given in Equations (3.5), we proceed to construct a nonparametric GMM series estimator by introducing a vector of instrumental variables $Z_{r,t}$. An example of $Z_{r,t}$ is to choose $Z_{r,t} = [\varsigma_1(X_t), \cdots, \varsigma_r(X_t)]'$ for some basis functions $\{\varsigma_j(x)\}_{j=1}^{r}$ which may differ from the basis functions $\{\phi\}_{j=1}^{p}$ and $\{\psi_j\}_{j=1}^{q}$. We require $2r \geq p + q = l$. Let $\otimes$ denote the Kronecker product. We obtain an extended set of orthogonality conditions defined as:

$$E[m^l(U_t, h)] = 0, \tag{3.15}$$

where $h \equiv \{f, g\}$, $U_t \equiv (X_t', Z_t')'$, $m^l(U_t, h) \equiv Z_{r,t} \otimes e_t(f, g) =$

$[m_1(U_t,h),\cdots,m_l(U_t,h)]'$ is a $l\times 1$ vector, and

$$e_t(h)=\begin{bmatrix}\beta^{\theta}(\frac{C_{t+1}}{C_t})^{\theta-1-\frac{\theta}{\eta}}(\frac{D_{t+1}}{D_t})(\frac{e^{g(X_{t+1})}}{e^{g(X_t)}-1})^{\theta-1}(e^{f(X_{t+1})}+1)-e^{f(X_t)}\\ \beta^{\theta}(\frac{C_{t+1}}{C_t})^{\theta-\frac{\theta}{\eta}}[\frac{e^{g(X_{t+1})}}{e^{g(X_t)}-1}]^{\theta}-1\end{bmatrix} \tag{3.16}$$

is the stochastic aggregate pricing error. The error terms are allowed to be conditionally heteroskedastic as well as serially correlated. We further denote a $l\times 1$ vector $m^l(h)=[m_1(h),\cdots,m_l(h)]'$, where $m_k\equiv E[m_k(U_t,h^o)]=E[m_{t,k}(h^o)]$ for all $k=1,\cdots,l$. Before introducing the definition of the nonparametric GMM series estimator for h^o, we first provide an identification condition that ensures the existence of a unique solution in a compact space.

Assumption 3.2.3. *There exists an unique $h^o\equiv(f^o,g^o)\in int(\Theta)$ such that $m^l(h^o)=0$ for all $l\geq 1$.*

Assumption 3.2.3 is an identification condition in a compact space. It ensures the existence of a unique solution h^o. Under the moment condition $m(h^o)=0$, h^o can be considered as the true solution of the unknown price-dividend ratio and wealth-consumption ratio functions. Let $\gamma^l=(a^{p'},b^{q'})'$ denote the unknown coeffcients, where $h_l(x)=[\phi^p(x)'a^p,\psi^q(x)'b^q]'$. We now define a nonparametric GMM series estimator.

Definition 3.2.1. *The nonparametric GMM series estimator $\hat{h}_l\equiv(\hat{f}_p,\hat{g}_q)$ is*

$$\hat{h}_l=arg\min_{h_l\in\Theta}\hat{Q}_T[h_l,\hat{W}(h_l)]=arg\min_{h_l\in\Theta_l}\frac{1}{2}\hat{m}_T'(h_l)\hat{W}^{-1}(h_l)\hat{m}_T(h_l). \tag{3.17}$$

where

$$\hat{m}_T(h_l)\equiv\frac{1}{T}\sum_{t=1}^{T}m^l(U_t,h_l)=[\hat{m}_{T1}(h_l),\cdots,\hat{m}_{Tl}(h_l)]',$$

is a $l \times 1$ *sample moment vector,* $\hat{m}_{Tj} \equiv \frac{1}{T}\sum_{t=1}^{T} m_j(U_t, h_l)$ *for* $j \in \{1, \cdots, l\}$*,* $\hat{W}(h_l)$ *is a* $l \times l$ *symmetric nonsingular matrix which is possibly data-dependent and parameter dependent, and* $h_l = [f_p, g_q]$ *contains a* $(p+q) \times 1$ *unknown parameter vector* γ^l*, and* Θ *is a 2-dimensional compact function space. Here, we assume* $l \geq p + q$*, i.e., the number of moments is larger than or at least equal to the number of parameters* γ^l*.*

Unlike Cui and Hong (2016), who consider a single Euler equation, there are generally no closed-form estimation solutions for multiple Euler equations in Equation (3.17). When $\hat{W} = I$, an identity matrix, each of the l component sample moments is weighted equality. If $\hat{W} \neq I$, then the l sample moment components are weighted differently. A suitable choice of weighting matrix $\hat{W}$ can improve the efficiency of the resulting series estimator. Based on whether the weighting matrix $\hat{W}(h)$ depends on unknown functions h, we consider two specific nonparametric GMM series estimator, namely the two-stage GMM series estimator and continuously updating efficient (CUE) GMM series estimator. Using the martingale difference sequence property of Equations (3.15), we obtain an optimal weighting matrix for the nonparametric GMM series estimator $W(h) = TE[\hat{m}_T(h)\hat{m}'_T(h)]$.

Utilizing the result that a consistent estimate of model parameters may be computed by GMM with an arbitrary positive definite and symmetric weighting matrix $\hat{W}$ such that $||\hat{W} - W|| = o_p(1)$, we introduce a two-step efficient nonparametric GMM series estimator. The weighting matrix $\hat{W}$ chosen for the two-stage efficient GMM series estimator does not involve unknown h_l. Therefore, after obtaining a preliminary consistent GMM series estimator $\tilde{h}_l$ with a prespecified weighting matrix $\tilde{W}$, say $\tilde{W} = I$, we can find a consistent variance-covariance estimator $\hat{W} = \frac{1}{T}\sum_{t=1}^{T} m(U_t, \tilde{h}_l)m(U_t, \tilde{h}_l)'$, which is further used in obtaining an asymptotically efficient estimator for h^o, namely the two-stage GMM series estimator. Specifically, because the weighting function $\hat{W}$ is constant with respect to unknown functions $h_l(x) = [\phi^p(x)'a^p, \psi^q(x)'b^q]'$, the first order condition with respect to $\gamma^l = (a^{p'}, b^{q'})' = (\gamma_1, \cdots, \gamma_l)'$ is

$$\frac{\partial \hat{Q}_T(h_l)}{\gamma_j} = \frac{\partial \hat{m}_T(h_l)'}{\partial \gamma_j} \hat{W}^{-1}(h_l)\hat{m}_T(h_l) = 0, \quad j = 1, \cdots, l. \tag{3.18}$$

We now propose the continuously updating efficient (CUE) nonparametric GMM series estimator, which has a weighting matrix $\hat{W}(h_l) = \frac{1}{T}\sum_{t=1}^{T} m(U_t, h_l)m(U_t, h_l)'$ depending on unknown functions h_l. The weighting function is continuously changed as h is altered in the minimization. Hansen et al. (1996) convince that the CUE GMM estimation of finite-dimensional parameters is invariant to how the moment conditions are scaled even when parameter-dependent scale factors are introduced. Using several different specifications of the CAPMs, Hansen et al. (1996) find that the finite-sample properties of the two-stage efficient GMM and the CUE GMM estimators of finite-dimensional parameters differ in the way in which the moment conditions are weighted. They provide evidence that it should be of use in many GMM estimation environment. In the exact CAPM context, Stock and Wright (1996) further point out that instruments may be only weakly correlated with the Euler equation errors thereby resulting in poor performance in normal approximations even in large samples. They posit that the CUE GMM provides better normal approximation in the finite sample than other GMM implementations. With general simultaneous moment conditions, Newey (2004) provide a theoretical rationale for considering the CUE GMM in situations in which models are weakly identified. In such settings, the accuracy of asymptotic approximations is largely improved by accounting for many moments. There is a rich literature that explores the accuracy improvement when formulating inference using the CUE GMM estimators for finite-dimensional parameters (e.g., Donald and Newey, 2000; Newey and Smith, 2004; Han and Phillips, 2006). Providing that asymptotic biases can be corrected using Jacobian matrix, Newey and Smith (2004) further explain the superiority of the finite dimensional CUE GMM estimators.

In an effort to improve the small sample properties of the nonparametric GMM series estimators, we consider a nonparametric CUE GMM procedure which can eliminate an important source of bias for GMM in models with endogeneity, while enhancing model inference for most DSGE setups. Allowing the weighting matrix to vary with unknown function h_l results in different first order conditions from the minimization. Specifically, partial derivatives with respect to γ_j for $j \in \{1, \cdots, l\}$ will be taken for both the weighting matrix $\hat{W}(h_l)$ and sample moments, which leads to the following moment conditions:

$$0 = \frac{\partial \hat{Q}_T(\psi'\hat{\gamma})}{\partial \gamma_j}, \quad j = 1, \cdots, l, \tag{3.19}$$

where

$$\frac{\partial \hat{Q}_T(h_l)}{\gamma_j} = \{\frac{\partial \hat{m}_T(h_l)'}{\partial \gamma_j}\hat{W}^{-1}(h_l)\hat{m}_T(h_l) - \hat{m}_T(h_l)'\hat{W}^{-1}(h_l)\frac{\partial \hat{W}(h_l)}{\partial \gamma_j}\hat{W}^{-1}(h_l)\hat{m}_T(h_l)\}. \tag{3.20}$$

Overall, we can find an optimal nonparametric GMM series coefficient estimator $\hat{\gamma}^l = [\hat{a}^{p\prime}, \hat{b}^{q\prime}]$ using either methods. It follows that the GMM series estimators of unknown functions f^o and g^o at any $x \in \mathbb{X}$ are given below

$$\begin{cases} \hat{f}_p(x) = \sum_{j=1}^{p} \phi_j(x)\hat{a}_j, \\ \hat{g}_q(x) = \sum_{j=1}^{q} \psi_j(x)\hat{b}_j. \end{cases} \tag{3.21}$$

For concreteness, we have introduced a nonparametric GMM series estimator in Epstein and Zin's (1989) model, which has two unknown functions in a system of two Euler equations. However, our approach can be extended in a straightforward manner to estimate more than two unknown functions in more complex DSGE models without many alternations. Considerable attempts to enrich the explanatory powers of economics models have been witnessed in recent years, thereby leading to increasing model complexity. For example, Liu et al. (2013) aim to capture the co-movement between land prices and macroeconomy by introducing the land-price dynamics in a DSGE model. As a result, there is an additional Euler equation that describes the optimal dynamics of land holding decisions. The nonparametric GMM series approach is applicable to this setup. It can provide estimation for all unknown functions simultaneously, which is asymptotically free of simultaneous equation biases, and accumulated functional form approximation errors. In the next section, we will establish the asymptotic properties of the proposed nonparametric GMM series estimators so as to facilitate statistical inference on model implications in practice.

3.3 Consistency and Asymptotic Normality

To investigate the asymptotic properties of the nonparametric GMM series estimator $\hat{h}_l$, we first provide a set of regularity conditions. Let $\lambda_{\min}(W)$ and $\lambda_{\max}(W)$ denote the minimum and maximum eigenvalues of the $l \times l$ weighting matrix W. Because we allow the dimension $l \equiv l(T) \to \infty$ as the sample size $T \to \infty$, we need to restrict the rate that $l \to \infty$ so as to ensure consistency.

Assumption 3.3.1. *(i) The weighting matrix* $||\hat{W} - W|| = o_p(1)$*, where* W *is a* $l \times l$ *symmetric, finite and nonsingular matrix; (ii)* $\lambda_{\min}(W) > 0$*; (iii)* $\lambda_{\max}(W) < \infty$*; (iv)* $\frac{lp^{-s}}{\lambda_{\min}(W)} \to 0$ *as* $T \to \infty$*; (iv) for all* $h \in \Theta$*,* $\lambda_{\min} W(h) = O(\lambda_{\min} W)$*.*

Assumption 3.3.1 (i) requires that a $l \times l$ weighting matrix $\hat{W}$ converge to W in probability as $T \to \infty$. When $\hat{W} = \frac{1}{T}\sum_{t=1}^{T} Z_t Z_t'$, Assumption 3.3.1 (i) is ensured by the condition that $\frac{l^2}{T} \to 0$ as $l \to \infty$ with $T \to \infty$, because $||\frac{1}{T}\sum_{t=1}^{T}(Z_t Z_t') - EZ_t Z_t'|| = O_p(\frac{l}{\sqrt{T}}) = o_p(1)$.

Assumptions 3.3.1 (ii) and (iii) impose some mild conditions on the behaviors of the maximum and minimum eigenvalues of W. Specifically, when $\hat{W} = \frac{1}{T}\sum_{t=1}^{T}(Z_t Z_t')$, Assumption 3.3.1 (ii) becomes the well-known necessary and sufficient condition for consistent estimation of parameters in a linear regression model with a fixed number of regressors (Drygas, 1976b). This assumption is employed by Andrews (1991) to establish consistency of the OLS series estimator when the number of regressors grows to infinity as $T \to \infty$. Cui and Hong (2016) impose this condition to establish consistency of a nonparametric 2SLS series estimator of the price-dividend ratio function in a single Euler equation. With a stronger assumption that $\lambda_{\min}[E(Z_t Z_t')]$ is uniformly bounded away from below from zero, Portnoy (1985) imposes Assumption 3.3.1 (iii) to obtain consistent estimation of parameters in a general linear regression model when the number of regressors tends to infinity as $T \to \infty$. As pointed out by Andrews (1991), Assumption 3.3.1 (ii) holds with probability one if $E(Z_t Z_t')$ is nonsingular for all $l \geq 1$. When this condition fails, basis functions that are redundant in the limit can be eliminated to ensure that Assumption 3.3.1 always holds. Assumption 3.3.1 (iv) implies that $\lambda_{\min} W(h)$ is at most of order $\lambda_{\min} W$ for all $h \in \Theta$. Specifically, it implies that there is always some constant $M > 0$ such that $[\lambda_{\min} W]^{-1} \lambda_{\min} W(h) \leq M$.

Assumption 3.3.2. *Let* $\{U_t\}_{t=1}^{T} \equiv \{X_t', Z_t'\}'^{T}_{t=1}$ *be a* $\beta-$*mixing process with mixing coefficients satisfy* $\sum_{j=1}^{\infty} j^2 \beta^{\frac{\delta}{1+\delta}}(j) < \Delta < \infty$ *for some* $0 < \delta < 1$.

Assumption 3.3.2 imposes a mild condition on the temporal dependence of state variables. This condition is common in the nonparametric literature (e.g., Chen and Hong, 2012). It indicates that the proposed nonparametric GMM series estimator is applicable to a rather wide class of stationary Markov processes. Put $F_p = e^{f_p}$ and $G_q = e^{g_q}$.

Assumption 3.3.3. *Let* (Θ, ρ) *be a metric space with* $\rho = ||\cdot||_{sd}$. $\cup_{l=1}^{\infty} \Theta_l$ *is dense in* Θ *in the metric* ρ. *For some* $0 < \Delta < \infty$, $t = 1, 2, \cdots, T$, *and* $j = 1, \cdots, l$, *(i)* $E|\varsigma_j(X_t)|^8 \leq \Delta < \infty$ *; (ii)* $E|e^{8X_t}| \leq \Delta < \infty$ *and* $E|e^{8(\theta - 1 - \frac{\theta}{\eta})X_t}| \leq \Delta < \infty$*; (iii)* $E|F_p|^8 \leq \Delta < \infty$, $E|G_q|^{8(1-\theta)} \leq \Delta < \infty$*; (iv)* $E|m_j^{4+\delta}(U_t, h)| \leq \Delta < \infty$ *for all* $h \in \Theta$*;*

The pseudo norm $||\cdot||_{sd}$ in Assumption 3.3.3 (i) can take various forms. It admits a range of convergence results with respect to different measures. In the present context, we consider the weighted sup-norm and Sobolev norms that $||\cdot||_{sd} = ||\cdot||_{d,w}$ for concretness. Assumptions 3.3.3 (ii, iii) impose some moment conditions on state variables and instrumental variables, which belong to the class of sub-exponential variables.

Theorem 3.3.1. *Suppose Assumptions 3.2.3-3.3.2 hold. Then for all* $t = 1, 2, \cdots, T$ *and* $h = 1, \cdot, l$*,*

(i) $m_k(U_t, h_l)$ *is a real-valued measurable function in* γ^l *;*

(ii) $m_k(U_t, h_l)$ *is Hölder continuous in* $\gamma^l \in \Theta_l$*, that is, there exists a constant* $\kappa \in (0, 1]$ *and a measurable function* $c_j(U_t)$ *with* $c_j(U_t) \in L^2_w$*, such that*

$$|m_j(U_t, t_1) - m_j(U_t, t_2)| \leq c_j(U_t)||\gamma_1 - \gamma_2||^{\kappa}_{sd}, \quad \textit{for all } X_t \in \mathbb{X}, \quad \gamma_1, \gamma_2 \in \Theta_l,$$

where $t_1(u) = [\phi^p(u)'{a_1}^p, \psi^q(u)'{b_1}^q]$*,* $\gamma_1 = [{a_1}^{p'}, {b_1}^{q'}]'$*,* $t_2(u) = [\phi^p(u)'{a_2}^p, \psi^q(u)'{b_2}^q]$*, and* $\gamma_1 = [{a_2}^{p'}, {b_2}^{q'}]'$*.*

Theorem 3.3.1 proves that each moment condition is stochastically equicontinuous. As pointed out by Newey (1991), stochastic equicontinuity of each sample moment is essential in establishing the property that $Q(h)$ is equicontinuous on Θ. It is also necessary for uniform convergence of $\hat{Q}_T(h_l, \hat{W})$ to $Q(h^o, W)$ in probability. With preliminary conditions on the behavior of state variables, we have actually proved that stochastic equicontinuity holds in the present context, without imposing it a high level assumption.

Theorem 3.3.2. *Suppose Assumptions 3.2.3-3.3.3 hold. Then as $l \to \infty$ with $T \to \infty$, we have $Q_T(h_l, \hat{W}) \xrightarrow{p} Q(h^o, W)$ uniformly for $h_l \in \Theta$.*

Theorem 3.3.2 ensures that $\hat{Q}_T(h_l, \hat{W})$ converges to $Q(h^o, W)$ in probability uniformly. It implies that when $\hat{h}_l$ maximizes the sample analog $\hat{Q}_T$, it will converge to the true unknown function h^o in probability as $T \to \infty$, where h^o uniquely maximizes $Q(h^o, W)$.

Theorem 3.3.3. *Suppose Assumptions 3.2.3-3.3.3 hold and $\hat{h}_l$ is a nonparametric GMM series estimator. Then for any given $x \in \mathbb{X}$, as $l \to \infty$ with $T \to \infty$, we have*

$$||\hat{h}_l(x) - h^o(x)||_{d,w} = O_p\big(\frac{l}{\lambda_{\min}(W)\sqrt{T}} + \kappa^{-s}\big).$$

Put $D_0 = \frac{dm^l(h_l)}{d\gamma}$, $G_{Tl} = D_0' W^{-1}(h) D_0$ and $\hat{G}_l = \frac{d\hat{m}^l}{d\gamma'} \hat{W}^{-1}(h) \frac{d\hat{m}^l}{d\gamma}$, all of which are $l \times l$ matrices. To establish the asymptotic properties of $a(\hat{h}_l)$, we need to estimate its sample variance-covariance.

Assumption 3.3.4. *Suppose there exists some $\Delta > 0$ such that for all $h_l \in \Theta_l$ and $i, j = 1, \cdots, l$, 1) $E|m_j(U_t, h_l)|^{4+\delta} \leq \Delta < \infty$; 2) $E|\frac{\partial m_j(U_t, h_l)}{\partial \gamma_i}|^{4+\delta} \leq \Delta < \infty$; 3) $E(\frac{\partial^2 m_j(U_t, h_l)}{\partial \gamma_i \partial \gamma_j})^{4+\delta} \leq \Delta < \infty$.*

The asymptotic normality for the series CUE GMM estimator differs from the results in Theorem 3.3.3. We need to impose an additional condition as follows.

Assumption 3.3.5. $\frac{l^4}{\lambda^6_{\min}(W)T} \to 0$ *as $l \to \infty$ with $T \to \infty$.*

This condition imposes a stronger restriction on the growth rate of the number of moments and the values of series truncation orders. The minimum eigenvalue $\lambda_{\min}(W)$ plays an important role here, since it also affects the rate at which the series truncation orders go to infinity. Without abuse of notations, we further denote

$$\hat{A}_j(h) = \frac{\partial \hat{W}(h)}{\partial \gamma_j}\hat{W}^{-1}(h),$$
$$A_j(h_{lo}) = \frac{\partial W(h_{lo})}{\partial \gamma_j}\hat{W}^{-1}(h_{lo}) = Tcov[\frac{\partial \hat{m}_T(h_{lo})}{\partial \gamma_j}, \hat{m}_T(h_{lo})]W^{-1},$$
$$\bar{U}^j(h) = \sqrt{T}\{\frac{\partial \hat{m}_T(h_{lo})}{\partial \gamma_j} - E[\frac{\partial \hat{m}_T(h_{lo})}{\partial \gamma_j}] - A_j\hat{m}_T(h_{lo})\}, \quad \bar{U} = [\bar{U}_1, \cdots, \bar{U}_l]',$$
$$U_t^j(h) = \{\frac{\partial m(U_t, h_{lo})}{\partial \gamma_j} - E[\frac{\partial m(U_t, h_{lo})}{\partial \gamma_j}] - A_j m(U_t, h_{lo})\}, j = 1, \cdots, l; t = 1, \cdots, T,$$
$$\Lambda_T = E[\bar{U}'W^{-1}\lambda_{\min}W\bar{U}]/l^2, \text{ and } \Lambda \equiv \lim_{T\to\infty}\Lambda_T.$$

Put $\hat{D}_j(h_{lo}) = \frac{\sqrt{T}}{l}[\frac{\partial \hat{m}_T(h_{lo})}{\partial \gamma_j}]$, $D(h) = E(\hat{D}(h))$, and $\hat{Q}_{\gamma\gamma'}(h_l) = \frac{T}{l^2}\frac{\partial^2 \hat{Q}_T(h_l)}{\partial \gamma_l \partial \gamma_l'}$. Also, put $H \equiv D(h^o)'W^{-1}\lambda_{\min}WD(h^o)$. A consistent estimator of H is $\hat{H} \equiv \hat{D}(h^o)'\hat{W}^{-1}\lambda_{\min}(\hat{W})\hat{D}(h^o)$.

Assumption 3.3.6. *Suppose there exists some $\Delta > 0$ such that for all $h_l \in \Theta_l$ and $i, j = 1, \cdots, l$, 1) $E(\frac{\partial m_j(U_t,h_l)}{\partial \gamma_i})^{2+2\delta} \le \Delta < \infty$; 2) $E(\frac{\partial^2 m_j(U_t,h_l)}{\partial \gamma_i \partial \gamma_j})^{2+2\delta} \le \Delta < \infty$.*

Theorem 3.3.4. *Suppose Assumptions 3.2.3-3.3.4 hold and $l \to \infty$ as $T \to \infty$. Then (i) $||\hat{H}(h_l) - H|| = o_p(1)$; (ii) $||\hat{Q}_{\gamma\gamma'}(h_l) - H|| = o_p(1)$.*

Theorem 3.3.5 is essentially helpful in obtaining consistent estimation of the variance matrix for the CUE GMM series estimator.

Theorem 3.3.5. *Suppose Assumptions 3.2.3-3.3.4 hold, and $\hat{h}_l$ is a two-stage GMM* *series estimator. For any given $x \in \mathbb{X}$, as $l \to \infty$ with $T \to \infty$,*

$$l\big(\hat{h}_l(x) - h^o(x)\big) \xrightarrow{d} N(0, H^{-1}). \tag{3.22}$$

To provide an intuition about the structure of the new variance-covariance matrix of the series CUE estimator, it is helpful to derive the following result,

$$l(\hat{h}_l(x) - h^{lo}(x)) = -\big[\frac{\partial^2 Q_T(\bar{h}_l)}{\partial\gamma\partial\gamma'}\lambda_{\min}(W)\frac{T}{l^2}\big]^{-1} l \frac{\partial Q_T(h_{lo})}{\partial\gamma}\lambda_{\min}(W)\frac{T}{l^2}$$

$$= -H(\bar{h}_l)^{-1}[D_j(h_{lo})'W^{-1}\lambda_{\min}(W)\sqrt{T}\hat{m}_T(h_{lo}) + \bar{U}'W^{-1}\lambda_{\min}(W)\sqrt{T}\hat{m}_T(h_{lo})/l] + o_p(1)$$

From Theorem 3.3.3, the first term in the bracket converges to a Gaussian distribution. The second term $\bar{U}'W^{-1}\lambda_{\min}(W)\sqrt{T}\hat{m}_T(h_{lo})/l$ is actually a degenerate U-statistic. We extend the results established by Gao and Hong (2007) and Chen and Hong (2010) for generalized U-statistic, and achieve asymptotic normality for this term when both the number of parameters and moments go to infinity as $T \to \infty$. We note that Donald and Newey (2000) and Newey (2004) establish the uncorrelatedness between these two terms.

Theorem 3.3.6. *Suppose Assumptions 3.2.3-3.3.4 hold, and $\hat{h}_l$ is a nonparametric CUE GMM series estimator. For any given $x \in \mathbb{X}$, as $l \to \infty$ with $T \to \infty$,*

$$l[\hat{h}_l(x) - h^o(x)] \xrightarrow{d} N(0, H^{-1} + H^{-1}\Lambda H^{-1}); \tag{3.23}$$

Theorem 3.3.6 indicates that there is an additional component in the variance-covariance matrix of the nonparametric CUE GMM series estimator. As shown in Theorem 3.3.5, this additional component H^{-1} is related to a higher (second) order derivatives, the Hessian matrix of $\hat{Q}_T$ with respect to γ^l. It is worth nothing that possible correlations among unknown functions which is captured by the variance terms is attributed to the interactions of unknown functions specified in the Euler equations (3.15). This correlation is important in practice. As pointed out by Newey (2004), when D_0 is close to zero relative to the variance of $\frac{m^l(U_t,h_l)}{\partial\gamma^l}$, it will lead to a much smaller $H^{-1}\Lambda H^{-1}$. Then the additional term H^{-1} will dominate even when $\frac{l}{T}$ is small. Therefore, correcting the variance term of the CUE type GMM series estimator is essential in obtaining reliable and meaningful inference. Then we obtain a heteroskedasticity-robust variance estimator $\hat{V}_{lT}$ for $\hat{f}_p(x)$:

$$
\begin{aligned}
\hat{V}_{lT} \equiv & [\hat{D}(h_{lo})'\hat{W}^{-1}\hat{D}(h_{lo})\lambda_{\min}(\hat{W}(h_{lo}))]^{-1} \\
& + [\hat{D}(h_{lo})'\hat{W}^{-1}\hat{D}(h_{lo})\lambda_{\min}(\hat{W}(h_{lo}))]^{-1}\hat{\Lambda}_T[\hat{D}(h_{lo})'\hat{W}^{-1}\hat{D}(h_{lo})\lambda_{\min}(\hat{W}(h_{lo}))]^{-1}.
\end{aligned}
$$

Theorem 3.3.7. *Suppose Assumptions 3.2.3-3.3.4 hold, and $\hat{h}_l$ is a nonparametric CUE GMM series estimator. Then for any given $x \in \mathbb{X}$, as $T \to \infty$, we have $\hat{V}_{lT} \xrightarrow{p} H^{-1} + H^{-1}\Lambda H^{-1}$.*

Theorem 3.3.7 implies that the sampling errors and approximation errors will not affect the consistency of the sample variance-covariance.

Compared with the nonparametric two-stage GMM series estimator, the variance of the nonparametric CUE GMM series estimator has an additional component H^{-1}, which captures the interaction between the two unknown functions within the system of Euler equations. This one-step estimation procedure significantly avoids accumulated approximation errors in comparison with all existing multi-step solution methods. We shall carry out three simulation studies in the next section to further illustrate more appealing features that our nonparametric GMM series estimators enjoy.

3.4 Empirical Applications and Simulation Studies

We now examine the finite sample performance of the nonparametric GMM series estimators. The first study is built on Tallarini's (2000) model, where the state variable X_t is a one-dimensional hidden Markov process. The second and third studies examine Bansal and Yaron's (2004) and Bansal et al.'s (2012) model, where two state variables are involved.

3.4.1 Tallarini's (2000) Model

We compare the finite sample performance of the nonparametric GMM series estimators with popular numerical solution methods in solving Tallarini's (2000) model. The state variable X_t is not directly observable, but it is embedded in a sequence of noise-driven observations, $c_{t+1} \equiv ln(\frac{C_{t+1}}{C_t})$ has the following dynamics:

$$\begin{cases} c_t = \mu t + X_t, \\ X_t = \Gamma X_{t-1} + \sigma_\epsilon \epsilon_t, \quad \text{where } \epsilon_t \sim N(0,1). \end{cases} \tag{3.24}$$

It is assumed that $C_t = D_t$ in each period. The univariate state variable $\{X_t\}_{t=0}^T$ summaries the law of motions in Tallarini's (2000) model. It is easy to conclude that

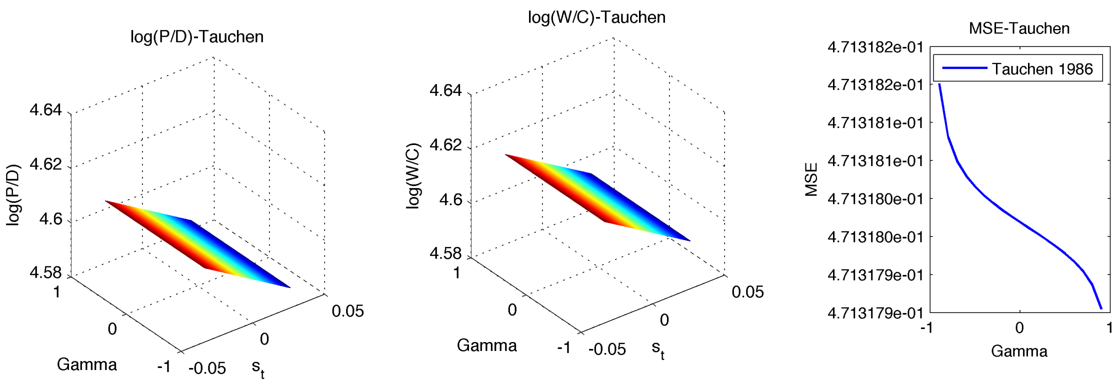
$ln(\frac{P_t}{D_t}) = f(X_t)$ and $ln(\frac{W_t}{C_t}) = g(X_t)$ in each period.

We consider four popular numerical solution methods for comparison, namely the log linearization, Tauchen's (1986) discretization method, Tauchen and Hussey's (1991) discretization method and the projection method. In practice, we determine the orders for f_t, g_t via the Akaike Information Criterion (AIC) for the nonparametric GMM series method.

We simulate a sample under Tallarini's (2000) DGP with 1,000 periods. The time discount factor $\beta = 0.99$, the relative risk aversion level $\gamma = 10$, the intertemporal elasticity of substitution (IES) $\eta = 1.5$, and the autocorrelation of the state variable Γ are equal to 0.91, and the unconditional mean of the state variable $\mu_s = 0$ with standard deviation $\sigma = 3.43\%$. The unconditional mean of the log consumption growth rate $\mu_c = 2\%$. Parameters like γ, Γ, and σ are allowed to vary at a reasonable range. $f_t \equiv ln(P_t/D_t)$, $g_t \equiv ln(W_t/C_t)$, and mean squared errors (MSE) are computed and plotted as a function of the state variable from different methods. In each group of graphs, the left panel depicts the dynamics of g_t, the middle panel plots f_t, and the right panel draws MSE. Three sets of graphs are presented in Figures 3.1-3.15. In the first set of graphs, as shown in Figures 3.1-3.5, results from the five solution methods are presented when the autocorrelation level Γ changes from -0.9 to 0.9. In the second set of graphs, as shown in Figures 3.6-3.10, γ varies from 1.05 to 20. The last set of graphs in Figures 3.11-3.15 shows the dynamics when the volatility σ of the state variable varies.

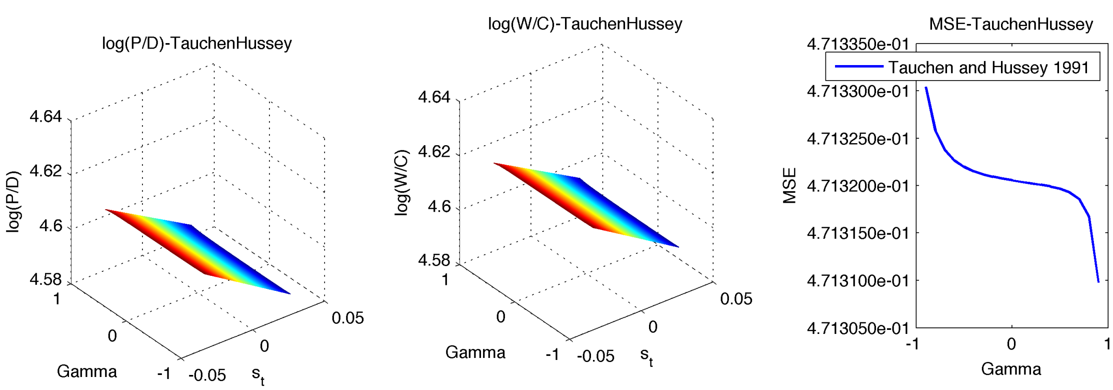
In the first set of analysis, both MSEs of two discretization methods increase

Figure 3.1: Price-dividend ratios, Wealth-consumption ratios and MSE as a function of Γ from Tauchen's (1986) Discretization Method



Notes: This figure provides approximated $ln(P/D)$ (left panel), $ln(W/C)$ (middle panel) and mean squared errors (MSE) from Tauchen's (1986) discretization method. As suggested by Tauchen (1986), the state variable s_t is transformed into a discrete space with N=9 states.

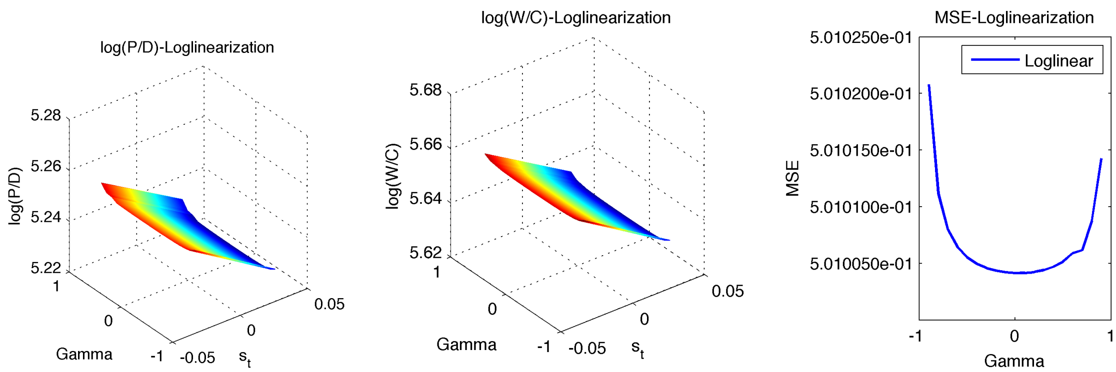
Figure 3.2: Price-dividend ratios, Wealth-consumption ratios and MSE as a function of Γ from Tauchen and Hussey's (1991) Discretization Method



Notes: This figure provides approximated $ln(P/D)$ (left panel), $ln(W/C)$ (middle panel) and mean squared errors (MSE) from Tauchen and Hussey's (1991) discretization method. As suggested by Tauchen and Hussey (1991), the state variable s_t is transformed into a discrete space with N=9 states.

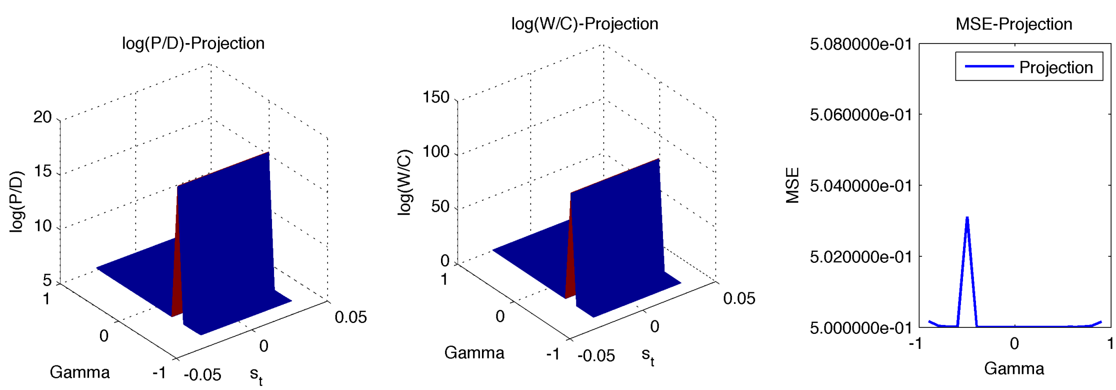
dramatically when Γ becomes large. It implies that the log linearization method is sensitive to the absolute value of Γ. The projection method is not severely affected by the value of Γ. All the log linearization, discretization and projection methods result in larger MSEs than the nonparametric GMM series estimator.

Figure 3.3: Price-dividend ratios, Wealth-consumption ratios and MSE as a function of Γ from Loglinearization Method



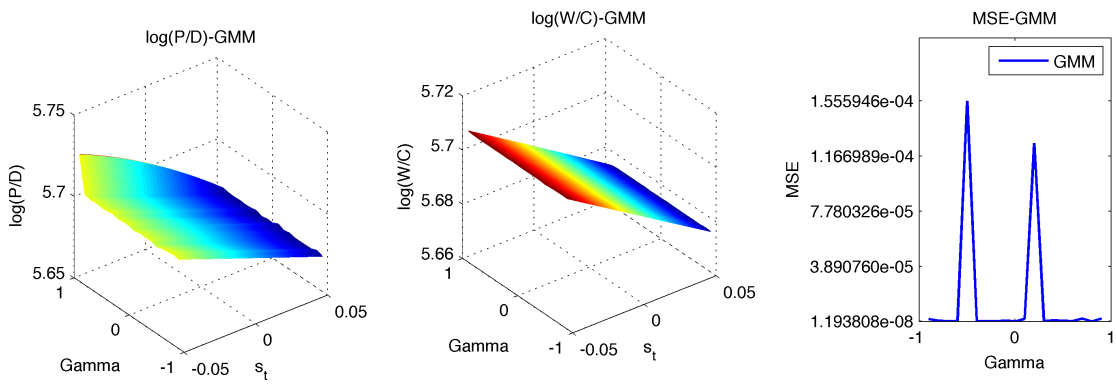
Notes: This figure provides approximated $ln(P/D)$ (left panel), $ln(W/C)$ (middle panel) and mean squared errors (MSE) from Loglinearization method.

Figure 3.4: Price-dividend ratios, Wealth-consumption ratios and MSE as a function of Γ from the Projection Method



Notes: This figure provides approximated $ln(P/D)$ (left panel), $ln(W/C)$ (middle panel) and mean squared errors (MSE) from the projection method. The order is chosen to be 2.

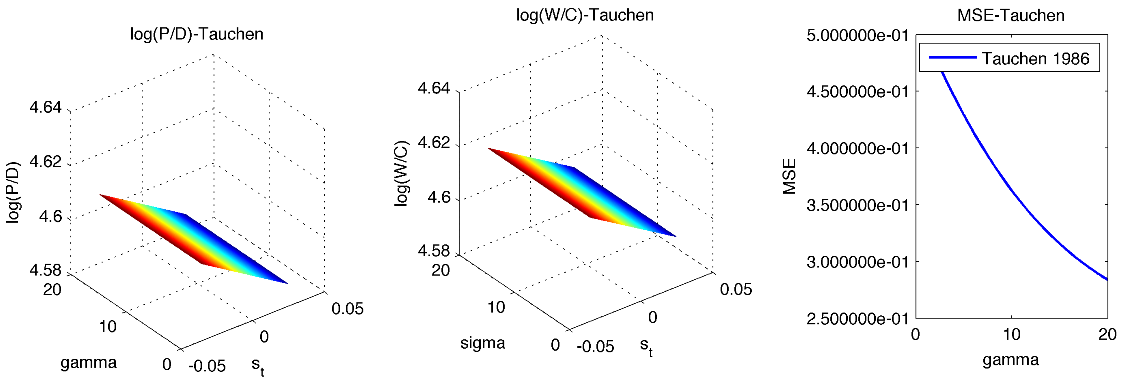
Figure 3.5: Price-dividend ratios, Wealth-consumption ratios and MSE as a function of Γ from the Nonparametric GMM Series Method



Notes: This figure provides approximated $ln(P/D)$ (left panel), $ln(W/C)$ (middle panel) and mean squared errors (MSE) from the GMM series estimation method. The order is chosen to be 2.

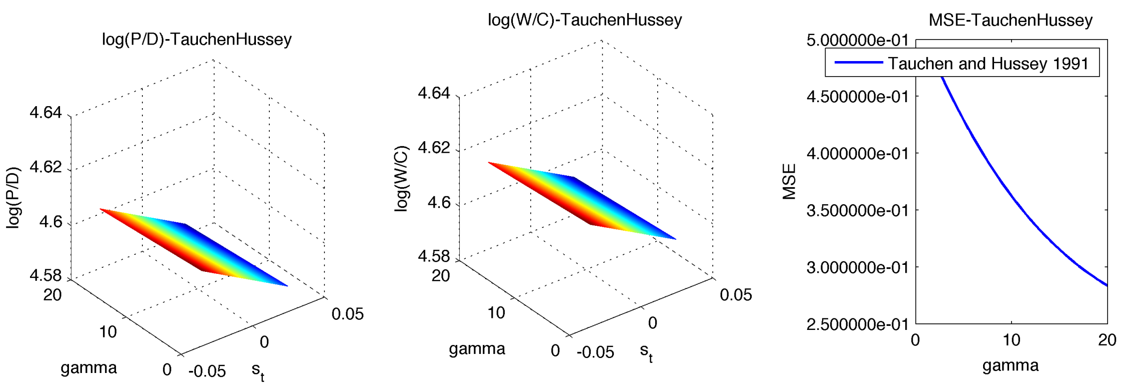
Figures 6-10 plot f_t and g_t when the risk aversion level γ changes. Starting from a mild risk-averse level, γ gradually increases to a highly risk-averse level. All the log linearization, discretization and projection methods exhibit monotonically increasing (or decreasing) capability in solution accuracy. Although the GMM series estimator has some variations in accuracy (apparently due to sampling variation in estimation), it achieves the smallest MSE and the best accuracy.

Figure 3.6: Price-dividend ratios, Wealth-consumption ratios and MSE as a function of γ from Tauchen's (1986) Discretization Method



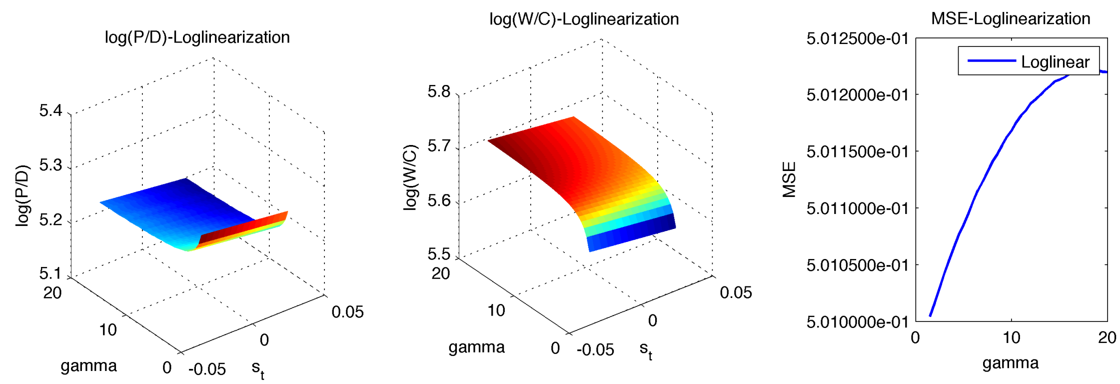
Notes: This figure provides approximated $ln(P/D)$ (left panel), $ln(W/C)$ (middle panel) and mean squared errors (MSE) from Tauchen's (1986) discretization method. As suggested by Tauchen (1986), the state variable s_t is transformed into a discrete space with N=9 states.

Figure 3.7: Price-dividend ratios, Wealth-consumption ratios and MSE as a function of γ from Tauchen and Hussey's (1991) Discretization Method



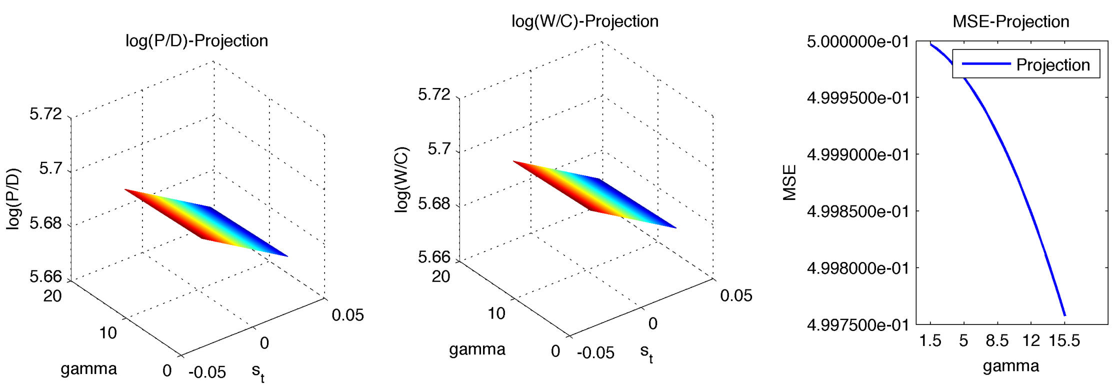
Notes: This figure provides approximated $ln(P/D)$ (left panel), $ln(W/C)$ (middle panel) and mean squared errors (MSE) from Tauchen and Hussey's (1991) discretization method. As suggested by Tauchen and Hussey (1991), the state variable s_t is transformed into a discrete space with N=9 states.

Figure 3.8: Price-dividend ratios, Wealth-consumption ratios and MSE as a function of γ from Loglinearization Method



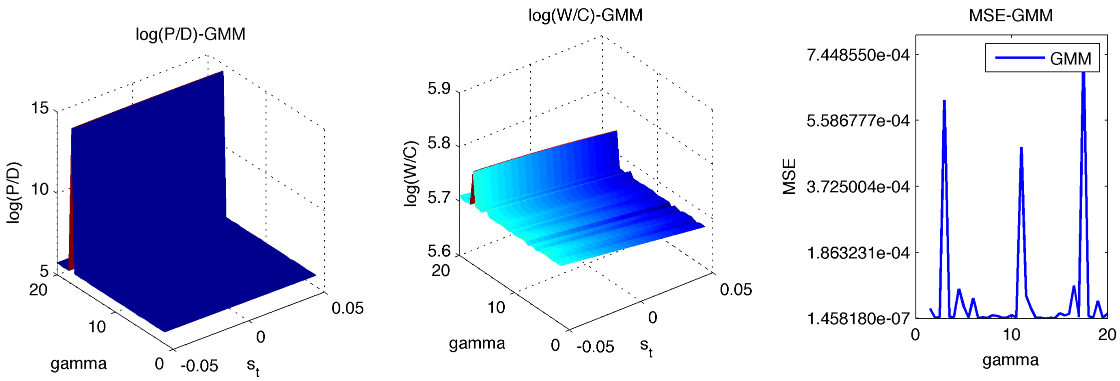
Notes: This figure provides approximated $ln(P/D)$ (left panel), $ln(W/C)$ (middle panel) and mean squared errors (MSE) from Loglinearization method.

Figure 3.9: Price-dividend ratios, Wealth-consumption ratios and MSE as a function of γ from the Projection Method



Notes: This figure provides approximated $ln(P/D)$ (left panel), $ln(W/C)$ (middle panel) and mean squared errors (MSE) from the projection method. The order is chosen to be 2.

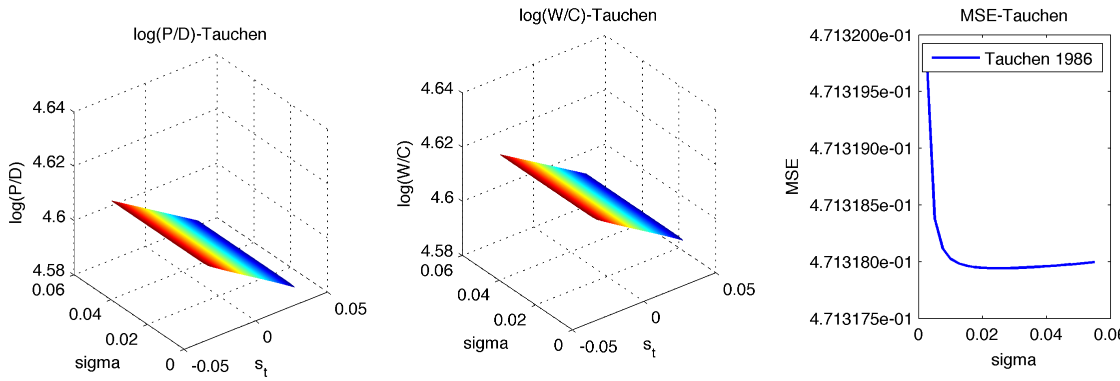
Figure 3.10: Price-dividend ratios, Wealth-consumption ratios and MSE as a function of Γ from the Nonparametric GMM Series Method



Notes: This figure provides approximated $ln(P/D)$ (left panel), $ln(W/C)$ (middle panel) and mean squared errors (MSE) from the GMM series estimation method. The order is chosen to be 2.

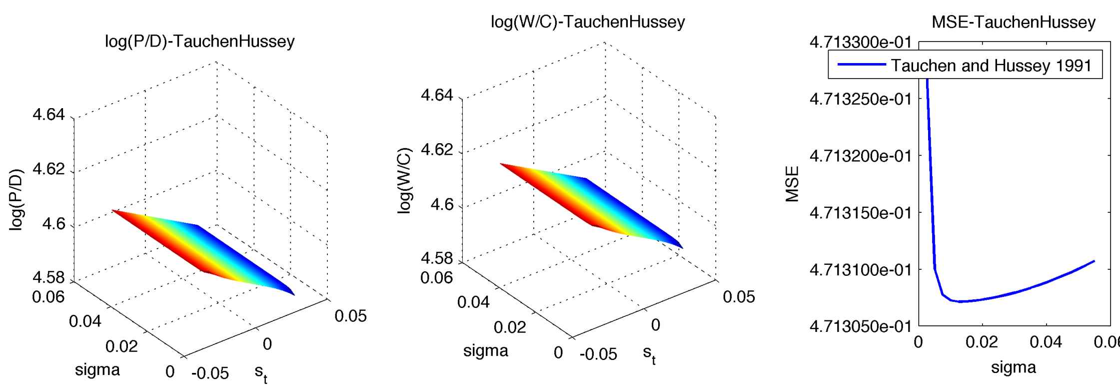
Last, we plot the logarithm price-dividend ratio and wealth-consumption ratio functions corresponding to different levels of the volatility level σ of the state variable. Although it has a high computational speed, the log linearization method suffers from exponentially increasing approximation errors when volatility increases. Two types of discretization methods perform unsatisfactory over their two tails. Again, the GMM series estimator turns out as the most reliable and accurate method in this situation.

Figure 3.11: Price-dividend ratios, Wealth-consumption ratios and MSE as a function of σ from Tauchen's (1986) Discretization Method



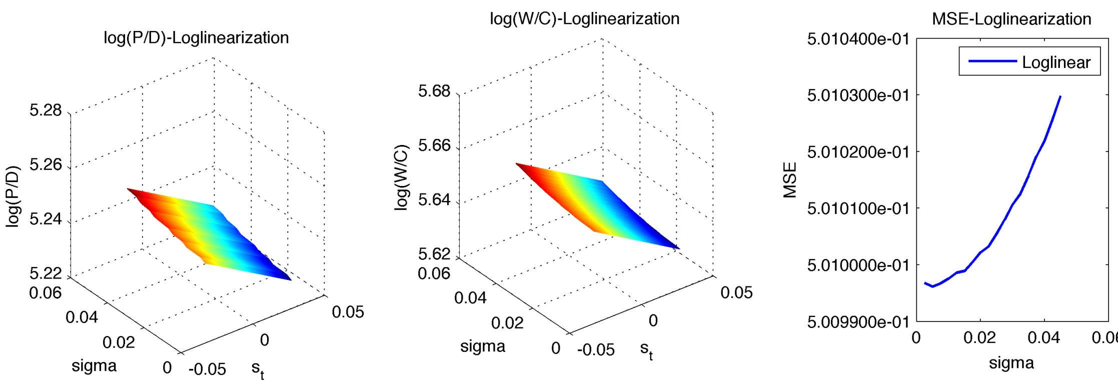
Notes: This figure provides approximated $ln(P/D)$ (left panel), $ln(W/C)$ (middle panel) and mean squared errors (MSE) from Tauchen's (1986) discretization method. As suggested by Tauchen (1986), the state variable s_t is transformed into a discrete space with N=9 states.

Figure 3.12: Price-dividend ratios, Wealth-consumption ratios and MSE as a function of σ from Tauchen and Hussey's (1991) Discretization Method



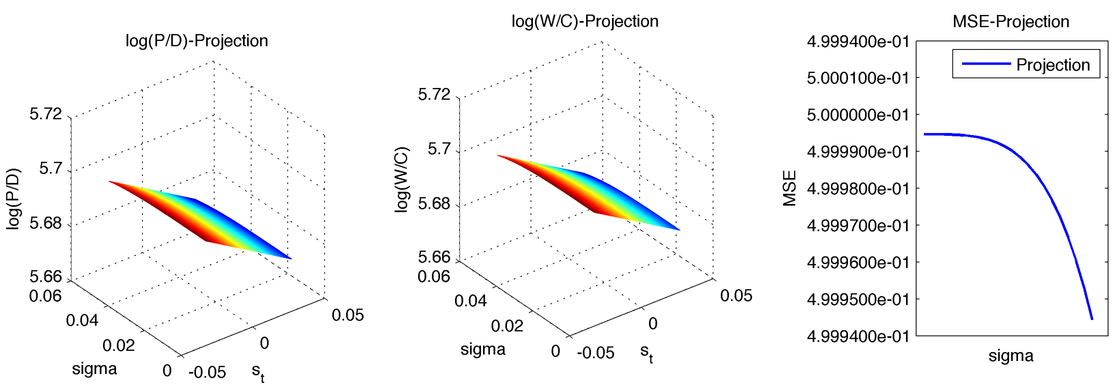
Notes: This figure provides approximated $ln(P/D)$ (left panel), $ln(W/C)$ (middle panel) and mean squared errors (MSE) from Tauchen and Hussey's (1991) discretization method. As suggested by Tauchen and Hussey (1991), the state variable s_t is transformed into a discrete space with N=9 states.

Figure 3.13: Price-dividend ratios, Wealth-consumption ratios and MSE as a function of σ from Loglinearization Method



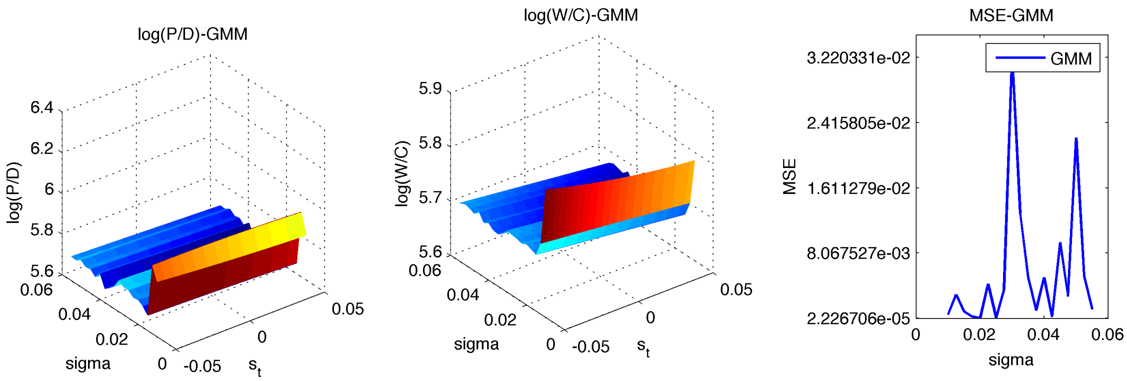
Notes: This figure provides approximated $ln(P/D)$ (left panel), $ln(W/C)$ (middle panel) and mean squared errors (MSE) from Loglinearization method.

Figure 3.14: Price-dividend ratios, Wealth-consumption ratios and MSE as a function of σ from the Projection Method



Notes: This figure provides approximated $ln(P/D)$ (left panel), $ln(W/C)$ (middle panel) and mean squared errors (MSE) from the projection method. The order is chosen to be 2.

Figure 3.15: Price-dividend ratios, Wealth-consumption ratios and MSE as a function of σ from the Nonparametric GMM Series Method



Notes: This figure provides approximated $ln(P/D)$ (left panel), $ln(W/C)$ (middle panel) and mean squared errors (MSE) from the GMM series estimation method. The order is chosen to be 2.

A central interest in the asset pricing literature is to understand the well-known equity premium puzzle. We find that the misspecification errors in approximating CAPMs will contaminate the results on equity premiums in a non-negligible manner. Table 3.1 reports the first two moments of the risky return, risk-free return, and price-dividend ratios from the empirical and simulation data from all four popular numerical solution methods and the nonparametric GMM series estimator. Although the discretization methods are able to capture the rough dynamics of the price-dividend ratio and wealth-consumption ratio functions, the interpolation biases are astounding when we use them to draw implications on asset returns. In Table 3.1, neither Tauchens (1986) algorithm nor Tauchen and Hussys (1991) algorithm can provide reasonable moments for the two assets. In the mean time, the log linearization method and the projection method launch different conclusions on equity premiums. Our nonparametric GMM series method reports that the average

Table 3.1: Asset Returns of the Tallarini's (2000) Model

	Empirical Data		Solution Methods				
Variables	Long Run	Post War	*LL*	*Tauchen*	*TH*	*Projection*	GMM series
$E(r_{f.t})$	1.9	1.7	2.29	0.91	0.91	2.79	1.89
$\sigma(r_{f,t})$	5.8	2.9	1.06	0.003	0.003	1.06	1.05
$E(r_{i,t+1} - r_{f,t})$	8.5	9.7	0.05	2.18	2.18	-0.33	0.54
$\sigma(r_{i,t+1})$	20.3	16.2	2.41	2.52	2.52	1.45	3.22
$E(p/d)$	3.2	3.4	5.21	4.59	4.59	5.69	5.89
$\sigma(p/d)$	0.4	0.5	$7e-4$	$1e-3$	$4e-4$	$2.5e-3$	$2e-4$

Note: All moments are in annual percentage. $R_{i,t+1}$ is the return of the risky asset. $R_{f,t}$ is the return of the risk-free asset. The long sample and postwar sample statistics are computed from the U.S. aggregate stock market. The long sample spans from 1890-2009 and the post-war sample spans from 1947-2009. The equity data are the Standard and Poor's 500 Price Index and Dividends. The risk-free rate is the return from the six-month commercial paper bought in January and rolled over in July. The U.S. aggregate stock market data are cited from http://www.econ.yale.edu/ shiller/data.htm. Simulations are conducted in annual frequency with sample size equal to 1000. $\beta = 0.99$, $\gamma = 10$, $\eta = 1.5$, $\Gamma = 0.91$, $\mu_s = 0$, $\mu_c = 2\%$ and $\sigma = 3.43\%$. This parametrization is adopted by Tallarini (2000).

equity premium is about 0.54% with a volatility level of 3.22%. From this simulation study, we find that model implications on asset returns differ dramatically from different solution methods. A confident and valid judgement on model evaluation should be built on a reliable and accurate solution method, and we recommend the proposed GMM series method due to its appealing asymptotic properties and reasonable finite-sample performance.

3.4.2 Bansal and Yaron's (2004) and Bansal et al.'s (2012) Models

For the second and third studies, we consider the Bansal and Yaron's (2004) and Bansal et al.'s (2012) long run risk models. These model have a long run predictable component and fluctuating economic uncertainty in conjecture with Epstein and Zin's (1989) preferences. These models are considered powerful in justifying equity premiums and enhancing understanding on economics anomalies.

Bansal and Yaron (2004) model fluctuating economic uncertainty as follows:

$$\begin{cases} \Delta c_{t+1} = \mu_c + X_t + \sigma_t u_{t+1}, \\ X_{t+1} = \Gamma X_t + \phi_e \sigma_t e_{t+1}, \\ \sigma_{t+1}^2 = \bar{\sigma}^2(1-\nu) + \nu\sigma_t^2 + \sigma_w w_{t+1}, \\ \Delta d_{t+1} = \mu_d + \phi X_t + \phi_d \sigma_t v_{t+1} + \pi \delta_t u_{t+1}, \end{cases} \tag{3.25}$$

where u_{t+1}, e_{t+1}, w_{t+1} and $v_{t+1} \sim i.i.d.N(0,1)$.

There are two state variables in each period that $S_t = (X_t, \sigma_t^2)$, namely the long run component X_t and the volatility level σ_t^2. The price-dividend ratio function is $f_t = f(S_t)$ and the wealth-consumption ratio function is $g_t = g(S_t)$. We solve this model under two specific parametrizations suggested by Bansal and Yaron (2004) and Bansal et al. (2012). Specific values of parametrizations are presented in Table 3.2. In practice, we decide the series truncation orders of f_t and g_t using AIC.

In reporting our simulation results, we use graphical methods. Figures 3.16-

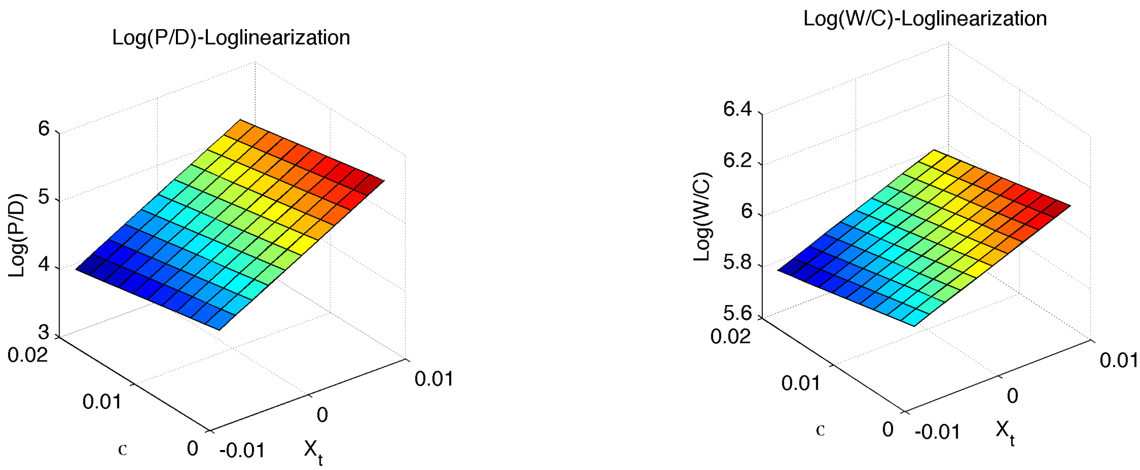
Table 3.2: Parametrizations of the Long-run Risk Models

	Preferences			DGPs									
DGPs	β	γ	η	μ_c	μ_d	Γ	ϕ	ϕ_d	ϕ_e	$\bar{\sigma}$	ν	σ_w	π
BY (2004)	0.9980	10	1.5	0.15%	0.15%	0.979	2.5	5.96	3.8%	0.72%	0.999	$2.8e-6$	2.6
BKY (2016)	0.9989	10	1.5	0.15%	0.15%	0.975	3.0	4.5	4.4%	0.78%	0.987	$2.3e-6$	0

Note: The first parametrization is adopted by Banal and Yaron (2004). The second set of parameters are used by Bansal et al. (2012)

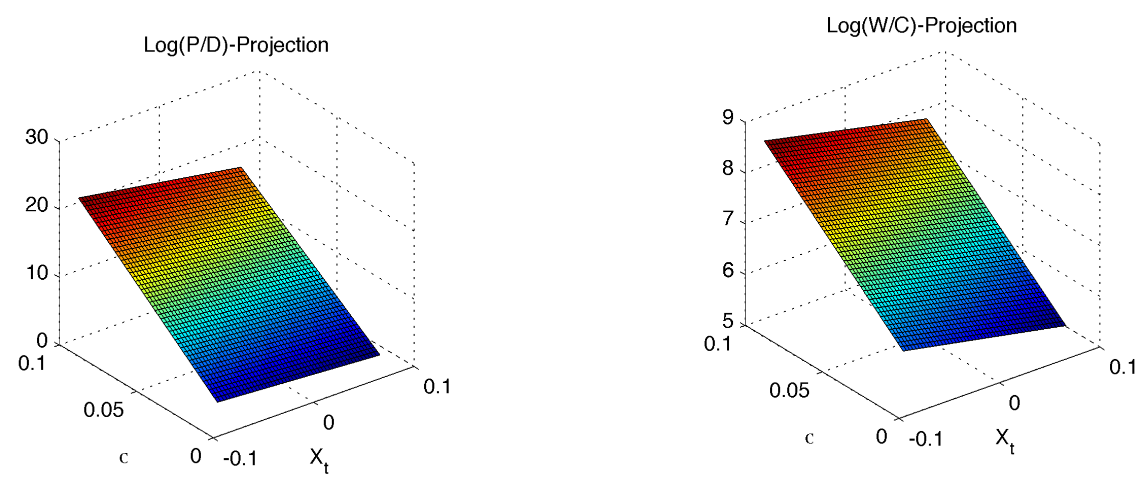
3.18 track the dynamics of the price-dividend ratio function f_t and the wealth-consumption ratio function g_t when X_t and σ_t^2 follow the law of motions described in Bansal and Yaron's (2004). We find that the approximated f_t and g_t are both increasing with σ_t^2 from the log linearization method, whereas they are shown to be negatively correlated with σ_t^2 from the projection method. The nonparametric GMM series method also admits a negative correlation between (f_t, g_t) and σ_t^2, whereas its estimation of (f_t, g_t) is smaller than that of the projection method. Similar results are seen in the third simulation study, where parameters values are set as in Bansal et al. (2012), which are reported in Figures 3.19-3.21. The resulting approximations from the log linearization method contradicts with that of the projection method and our GMM series method. This is mainly due to ignoring higher order approximations. While being fast and reasonably under near logarithm utility functions, the log linearization method has substantial misspecification errors as are captured by MSE.

Figure 3.16: Price-dividend ratios, Wealth-consumption ratios from Loglinearization method under Bansal and Yaron's (2004) parametrization



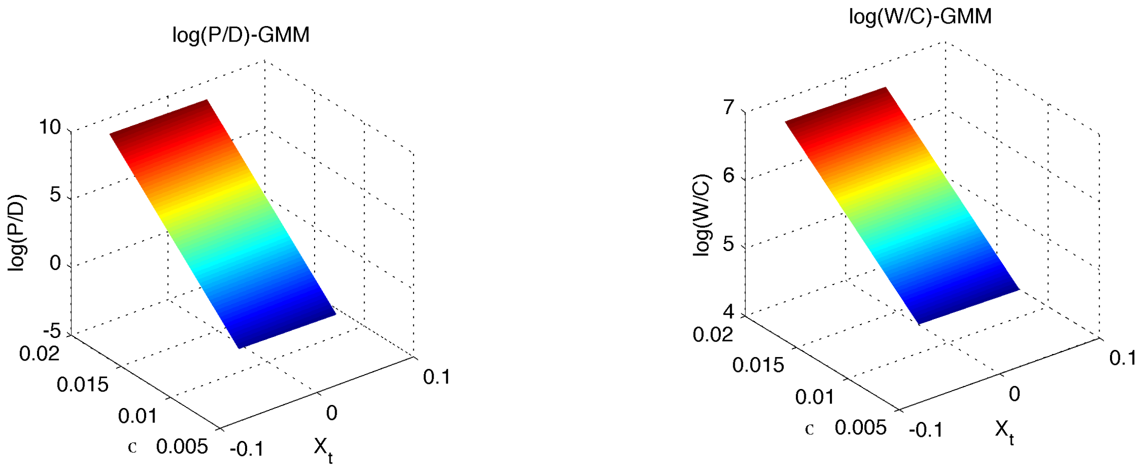
Notes: This figure provides approximated $ln(P/D)$ (left panel), $ln(W/C)$ (right panel) from the Loglinearization method. The state variable $s_t = [x_t, \sigma_t]$ is transformed into a discrete space with N=9 states.

Figure 3.17: Price-dividend ratios, Wealth-consumption ratios from the projection method under Bansal, Kiku and Yaron's (2012) parametrization



Notes: This figure provides approximated $ln(P/D)$ (left panel), $ln(W/C)$ (right panel) from the projection method. The state variable $s_t = [x_t, \sigma_t]$. We simulate a 1000 sample.

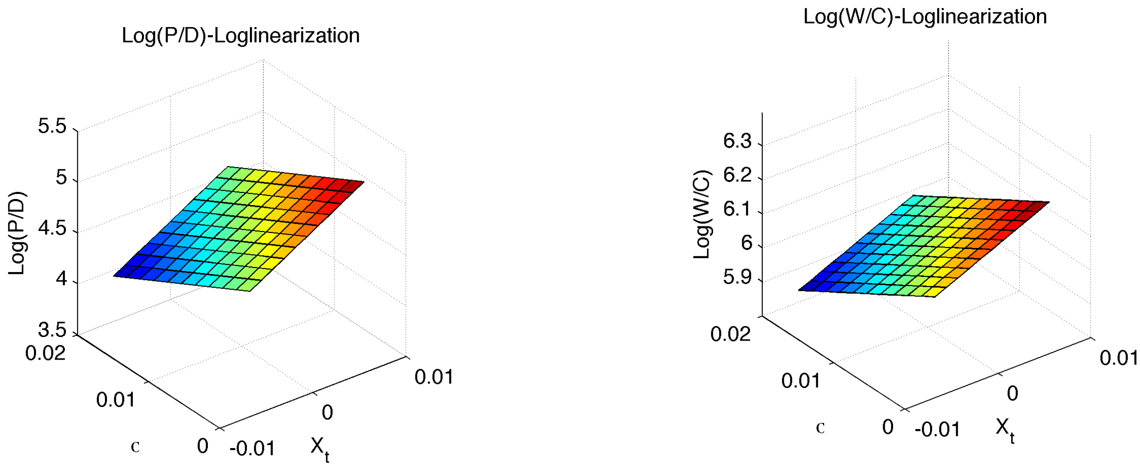
Figure 3.18: Price-dividend ratios, Wealth-consumption ratios from the Nonparametric GMM Series estimation method under Bansal, Kiku and Yaron's (2004) parametrization



Notes: This figure provides approximated $ln(P/D)$ (left panel), $ln(W/C)$ (right panel) from the projection estimation method with Garlerkin weighting functions. The state variable $s_t = [x_t, \sigma_t]$.

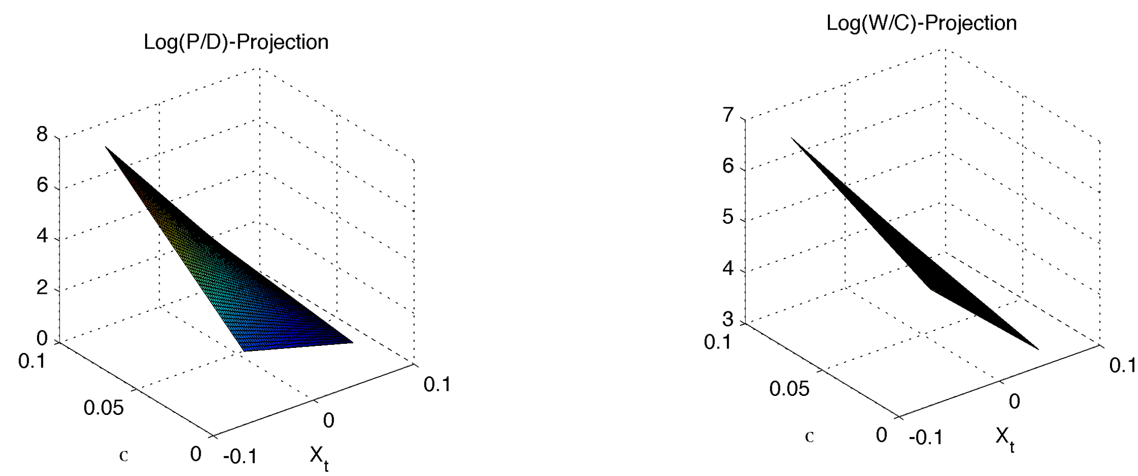
Table 3.3 provides model implied moments on asset returns under the log-linearization, projection and nonparametric GMM series methods. It is convincing that the introduction of a dividend process in Bansal et al.'s (2012) model is truly helpful in addressing equity premiums and other recorded anomalies to some extent. But due to functional form misspecification errors, existing numerical solution methods lead to different conclusions on equity premiums. We suggest the GMM series method in gaining more trustful model implications.

Figure 3.19: Price-dividend ratios, Wealth-consumption ratios from Loglinearization method under Bansal, Kiku and Yaron's (2012) parametrization



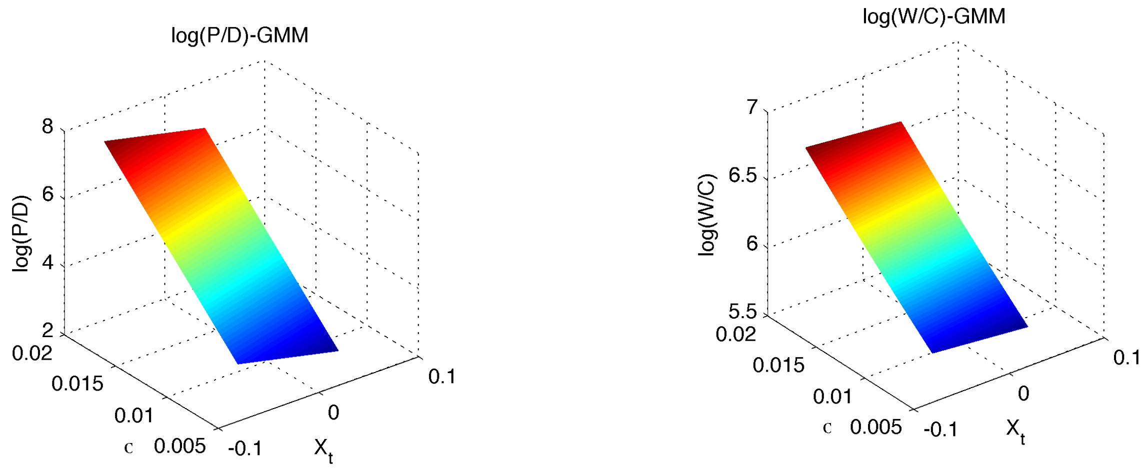
Notes: This figure provides approximated $ln(P/D)$ (left panel), $ln(W/C)$ (right panel) from the series Loglinearization method. The state variable $s_t = [x_t, \sigma_t]$ is transformed into a state space model with N=9.

Figure 3.20: Price-dividend ratios, Wealth-consumption ratios from the projection method under Bansal, Kiku and Yaron's (2012) parametrization



Notes: This figure provides approximated $ln(P/D)$ (left panel), $ln(W/C)$ (right panel) from the projection method. The state variable $s_t = [x_t, \sigma_t]$. We simulate a 1000 sample.

Figure 3.21: Price-dividend ratios, Wealth-consumption ratios from the Nonparametric GMM series estimation method under Bansal, Kiku and Yaron's (2012) parametrization



Notes: This figure provides approximated $ln(P/D)$ (left panel), $ln(W/C)$ (right panel) from the GMM series estimation method. The state variable $s_t = [x_t, \sigma_t]$. We simulate a 1000 sample.

3.5 Conclusion

DSGE models are considered as a pivotal tool in understanding financial markets and the macroeconomy. However, solving DSGE models accurately is a widely acknowledged challenge, because it involves a set of Euler equations, where there are multiple unknown functions. These unknown functions are interactive with each other and recursively specified under rational expectations. Analytic solutions become extremely difficult, if not impossible, especially for DSGE models with recursive preferences. Therefore, the current literature in finance and macroeconomics obtain model implications by undertaking some popular numerical solution methods,

Table 3.3: Asset Returns of the Long-Run Risk Models

Variables	Empirical Data: Long Run	Empirical Data: Post War	Solution Methods: *Loglinearization*	Solution Methods: *Projection*	Solution Methods: GMM series
Bansal and Yaron (2004)					
$E(r_{f,t})$	1.9	1.7	1.47	2.44	2.63
$\sigma(r_{f,t})$	5.8	2.9	0.40	0.40	0.39
$E(r_{m,t+1} - r_{f,t})$	8.5	9.7	3.61	6.04	2.04
$\sigma(r_{m,t+1})$	20.3	16.2	15.40	18.20	13.49
$E(p/d)$	3.2	3.4	5.05	5.03	5.43
$\sigma(p/d)$	0.4	0.5	0.14	0.19	0.09

Variables	Empirical Data: Long Run	Empirical Data: Post War	Solution Methods: *Loglinearization*	Solution Methods: *Projection*	Solution Methods: GMM series
Bansal, Kiku and Yaron (2012)					
$E(r_{f.t})$	1.9	1.7	2.69	1.66	1.67
$\sigma(r_{f,t})$	5.8	2.9	0.37	0.41	0.41
$E(r_{i,t+1} - r_{f,t})$	8.5	9.7	1.30	3.20	3.20
$\sigma(r_{m,t+1})$	20.3	16.2	17.36	19.49	19.49
$E(p/d)$	3.2	3.4	5.39	5.03	5.03
$\sigma(p/d)$	0.4	0.5	0.15	0.14	0.14

Note: All moments are in annual percentage. $R_{m,t+1}$ is the return of the risky asset. $r_{m,t+1}$ is logarithm of $R_{m,t+1}$. $R_{f,t}$ is the return of the risk-free asset. $r_{f,t}$ is the logarithm of $R_{f,t}$.The long sample and postwar sample statistics are computed from the U.S. aggregate stock market. The long sample spans from 1890-2009 and the postwar sample spans from 1947-2009. The equity data are the Standard and Poor's 500 Price Index and Dividends. The risk-free rate is the return from the six-month commercial paper bought in January and rolled over in July. The U.S. aggregate stock market data are cited from http://www.econ.yale.edu/ shiller/data.htm. Simulations are conducted in annual frequency with sample size equal to 1000. $\beta = 0.99$, $\gamma = 10$, $\eta = 1.5$, $\Gamma = 0.91$, $\mu_s = 0$, $\mu_c = 2\%$ and $\sigma = 3.43\%$. This parametrization is adopted by Banal and Yaron (2004). The Bansal, Kiku and Yaron's (2012) parametrization is the same as depicted in table 2.

namely the log-linearization, discretization and projection methods. Unfortunately, all existing numerical solution methods suffer from functional form misspecification errors for individual unknown function. They have to specify the full dynamics of state variables, which may also suffer from model misspecification.

To fill this gap, we introduce a nonparametric GMM series estimation method in the context of consumption-based asset pricing models with recursive preferences and multiple Euler equations. Our new method can estimate all unknown functions simultaneous, without involving sequential estimations. Moreover, this new method has been proven to be asymptotically free of functional form misspecification and simultaneous equation biases when the sample size increases. Unlike all existing numerical solution methods, it only assumes that state variables are Markov processes, and does not require any specification for the dynamics of state variables.

This paper proposes two types of the nonparametric GMM series estimators, namely the two-stage GMM series estimator and the CUE GMM series estimator. When Euler equations are weakly identified, the CUE type estimator can correct the variance of the series estimator, therefore provide a much reliable estimation on the price-dividend ratio and wealth-consumption ratio functions. Our empirically relevant simulation studies show that the nonparametric GMM series method performs substantially well in finite samples and under various parametrizations. It outperforms the existing numerical solution methods such as the log-linearization, discretization and projection methods.

This paper can be extended in several dimensions. First, it is applicable to DSGE models with more than two unknown functions without any alternations. Therefore, the solution accuracy can be significantly enhanced by eliminating all possible accumulated functional-form approximation errors occurred during the multi-step approximation procedures using all existing solution methods. Second, our method can be extended to DSGE models in the production economy, where log-linearization method is widely used for computational convenience (Zietz, 2006). It can help providing more reliable and accurate impulse functions for economic shocks, which can facilitate formulating effective policies in the real economy.

CHAPTER 4

EXTRAPOLATION BIAS IN ECONOMIC FUNDAMENTALS AND THE AGGREGATE STOCK MARKET

CAPM has been widely applied in theoretical finance and macroeconomics as a pivotal tool in understanding stock markets and macro economies. This model serves as a cornerstone in this strand of literature. However, despite the existence of this model, several economic anomalies are still difficult to explain, including the equity premium puzzles and accumulative excess returns. Significant efforts have been exerted in recent years to enrich the explanation power of CAPM by introducing empirically supported factors. In traditional finance, Eichenbaum et al. (1988) confirm the role played of by leisure time in understanding equity premium puzzles. Campbell and Cochrane (1999) discover that consumption habits help explain many economic phenomena in the stock market. Bansal and Yaron (2004) and Bansal et al. (2012) point out that long-run risks can significantly affect stock prices. Rietz (1988), Barro (2006), and Wachter (2013) consider rare disasters into asset pricing models. These papers have largely widened our understanding of pricing mechanisms from a theoretical perspective. This positive effect is attributed to the difficulty of obtaining enough observations for events such as rare disasters from the real economy. The empirical contributions of this work to policy makers are difficult to justify. In the last two decades, an increasing number of psychological evidence is discovered, and irrational factors are introduced into the asset pricing literature. For example, Barberis et al. (1999) explain the aggregate stock market by introducing prospect theory, which is a preference-biased approach. The concept of narrow framing in Barberis and Huang (2009), which is a local manner of viewing risks, also helps us understand many documented facts in finance. The theoretical models used by these papers to understand stock markets are CAPM-based or are derivatives of CAPM. We continue on this assumption and explore the modern stock markets in China and the United States using the CAPM-based model of Lucas Jr (1978).

The survey findings by Greenwood and Shleifer (2014) indicate that a large share of investors, including individuals, CFOs, and professional investors, hold extrapolative expectations. These investors act on their distorted beliefs and tend to forecast future stock prices using historical performance. Given that their work casts serious doubt on the full rationality assumption, which is popularly assumed in the current dynamic stochastic equilibrium models, there exists an urgent need to revisit possible irrationality in DSGE models and explore the role played by distorted beliefs in addressing those anomalous facts in the aggregate market. Barberis et al. (2015) introduce price extrapolation biases into a two-representative agent CAPM setup, which has a fully rational investor and a price extrapolator. After incorporating this kind of distorted beliefs, the present paper can capture the first two moments of the returns for the 6-month commercial bills and $S\&P$ 500. Barberis et al. (2015) theoretically improve extrapolative biases to enable them to be accepted and restudied. They also suggest that the survey evidence of Greenwood and Shleifer (2014) should be a key in understanding the aggregate market instead of treating it as an obstacle in the finance study.

Our paper discovers that investors from China and the United States hold different types of distorted beliefs on economic fundamentals. To the best of our knowledge, our paper is among the first to confirm and quantify this difference in the literature. Theoretically, we start from the CAPM setup of Lucas Jr (1978). In this setup, a representative agent holds extrapolative biases on economic fundamentals. By optimizing the agent's consumption stream, positions on a risk-free asset, and a risky asset throughout his life, he maximizes expected lifetime utility at time zero. Dividends from the risky asset are assumed to be i.i.d. normal, which is considered a golden rule in literature. However, given the limited time and knowledge, the agent may fail to capture this underlying distribution. In this sense, the distorted belief is about the entire distribution of economic fundamentals, which is not restricted to the status of the economy similar to the finding in Cecchetti et al. (2000). Our paper incorporates distortions in volatilities, and estimates how investors tend to extrapolate wrongly on volatile levels of economic fundamentals in their stock trading activities.

One difficulty faced by the current macroeconomics and finance literature is how to obtain correct estimations under non-model consistent frameworks. When investors hold distorted beliefs in model-driving factors, subjective expectations will differ from the mathematical one. However, the traditional GMM estimation only works under full rationality, where subjective estimation coincides with mathematical expectation. Our paper contributes to literature by proposing a new GMM estimation method that can work with subjective expectations. We consider extrapolation biases in volatilities along with autoregressive type mean level extrapolation biases.

Two sources of moment functions are employed in this paper. This first group of moment functions consists of differences between equity returns and simulated moments (McFadden, 1989). The second source of moment functions are constructed by the new GMM estimation procedure. Our paper quantifies and estimates specific extrapolation biases for Chinese and American investors. Using historical data from 20022015, we find that American investors hold extrapolation biases in the mean levels of economic fundamentals. Chinese investors, meanwhile, tend to ignore changes in the mean levels, but they focus on the overall volatile level of economic fundamentals. This finding sheds light on the phenomenon that positive but short-run policies can stimulate the stock market in the United States, but such policies fail to receive sufficient feedback from the stock market in China. Only sustainable and long-run policies can receive enough attention from the public. This situation improves investor confidence in China. Given that the majority of stock market participants in China are individual investors, correcting their distorted beliefs and enhancing their confidence of the economic background is the long-term objective of the Chinese government and stock market regulators.

4.1 Models

We first consider the traditional CAPM model of Lucas Jr (1978), which has one representative agent in an endowment economy. Let $\{C_t\}_{t=0}^{\infty}$ denote consumptions at time t, R_f denotes the return of risk-free asset at time t. P_t is the price of the risky asset at time t, which is a claim to future dividends D_t. We further denote

$z_t \equiv log(\frac{D_t}{D_{t-1}})$ as the logarithm of dividend ratios. The traditional CAPM model assumes the law of motion for z_t as

$$z_{t+1} = \mu_r + \epsilon_{t+1} \tag{4.1}$$

where $\epsilon_{t+1} \sim i.i.d.N(0, \delta_r^2)$.

The investor maximizes his expected lifetime utility at time zero by optimizing his consumption streams and shares of equities throughout his life:

$$\boldsymbol{E}[\sum_{t=0}^{\infty} \beta^t \frac{C_t^{1-\gamma}}{1-\gamma}] \tag{4.2}$$

where β is the time discount factor and γ is the relative risk-aversion level. Under the full rationality assumption, the representative agent is assumed to have the ability to know exactly the dynamics of driving factor z_t. This assumption means that the investors subjective expectation coincides with mathematical expectation. In equilibrium, optimal consumption at each period is equal to the dividend payment in that period $D_t = C_t$. This equation serves as the theoretical reason for choosing consumption growth rates as an index for economic fundamentals in our paper. For ease of notation, we define the pricedividend ratio $\omega_t = \frac{P_t}{D_t}$ and rewrite the Euler equation as

$$\begin{aligned} \omega_t &= \boldsymbol{E}[\beta(\frac{C_{t+1}}{C_t})^{1-\gamma}(\omega_{t+1}+1)|\boldsymbol{I_t}] \\ &= \boldsymbol{E}[\beta(e^{(1-\gamma)(\mu_r+\epsilon_{t+1})}(\omega_{t+1}+1)|\boldsymbol{I_t}]. \end{aligned} \tag{4.3}$$

where $\boldsymbol{I_t}$ denotes all available information until time t. The return of the risky asset is given by

$$R_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t} = \frac{\frac{P_{t+1}}{D_{t+1}} + 1}{\frac{P_t}{D_t}} \frac{D_{t+1}}{D_t}$$
$$= \frac{\frac{P_{t+1}}{D_{t+1}} + 1}{\frac{P_t}{D_t}} e^{\mu_r + \epsilon_{t+1}} \tag{4.4}$$
$$= \frac{\omega_{t+1} + 1}{\omega_t} e^{\mu_r + \epsilon_{t+1}}$$

The return of the risk-free asset is given by

$$R_f = \frac{1}{\beta} e^{\gamma\mu_r - \frac{\gamma^2}{2}\delta_r^2} \tag{4.5}$$

Despite its theoretical contribution in the literature, this model fails to address economic anomalies. One of the biggest challenges of this fully rational CAPM equation is the equity premium puzzle, first pointed out by Mehra and Prescott (1985). We keep everything rational in all aspects as we enhance model performance by introducing two kinds of possible distorted beliefs in consumption growth rate.

4.1.1 Models with Extrapolation Biases in Economic Fundamentals

The underlying economic fundamental in models with extrapolation biases is assumed to follow Equation 4.1. Given the limitations in time, ability, and other factors, investors fail to grasp correctly this law of motion. They may use historical data to help them produce forecasts on the dynamics of economic fundamentals.

Two kinds of distorted beliefs are studied in this paper, namely, extrapolation biases in mean and in variance.

The distribution of economic fundamentals is assumed to be i.i.d., which is considered a golden benchmark in related literature. We retain the utility function ordinary and the true underlying process typical. In this approach, the only driving factor that differentiates our model from others is distorted beliefs. This approach helps purify the influence of extrapolation biases on understanding the mechanism of pricing schemes. In reality, a limited number of observations exist since the birth of stock markets in all nations. The hidden operation mechanism of economic fundamentals is difficult to determine know. Therefore, we need to ascertain the validity of normal distribution assumption rigorously. We employ the one-sample Kolmogorov-Smirnov test on the consumption growth rate data of China and the United States from 20022015.

Table 4.1: Kolmogorov-Smirnov Test P-value

	China	U.S.	Japan	U.K.
P-value	0.568	0.347	0.346	0.488

We use the P-value of the Kolmogorov-Smirnov test to judge the assumption on normal distribution. Table 1 shows that the P-value of the operating mechanism in the United States, Japan and U.K. are 0.347, 0.346 and 0.488, respectively. It implies that the assumption on the normal distribution for the economic fundamentals is acceptable under the 1%, 5%, and 10% significance levels. These values are strong pieces of evidence on the normal distribution assumption.

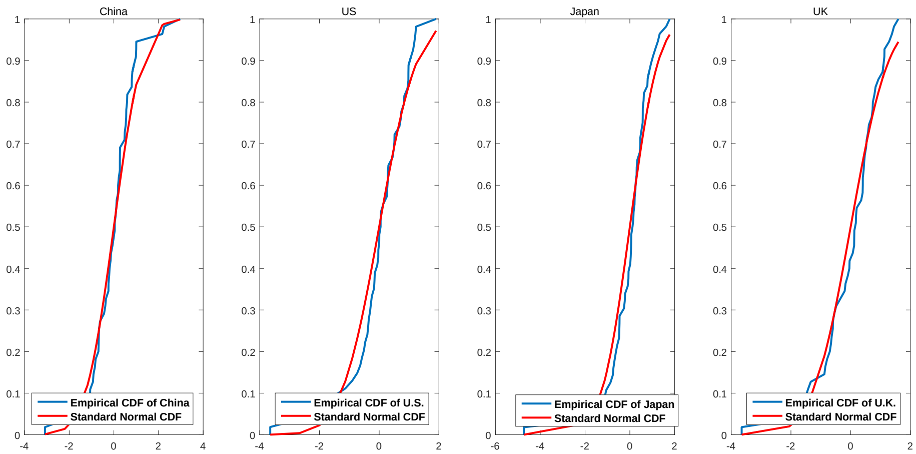


Figure 4.1: Kolmogorov-Smirnov test

Whenever there are large amounts of observations, all economic models will be rejected by rigorous statistical tests. Despite this fact, we need to choose the most helpful economic model to help us perform instructive analysis on economic phenomena Campbell and Cochrane (2000). This finding is attributed to the fact that we can take advantage of these models to discuss the possible driving forces of the model results. Given the short sample periods and limited market information, we consider the normal distribution assumption as acceptable assumption, which we use in the rest of our paper. A large amount of work in this strand of literature continues to rely on the normal distribution assumption (Barberis et al., 2015), which provides us a fair ground upon which to discuss model performance with the most related work.

Model 1: Investor Extrapolation Biases in the Mean Level of Economic Fundamentals

To differentiate the objective operating process and the investors subjective expectation, we let E^s denote the distorted beliefs held by the investor. Barberis et al. (2015) suppose that in forecasting future stock performance, the investor will extrapolate on the entire historical observations with an exponentially decreasing weight. This assumption on distorted belief is too strong and lacks empirical evidence. Such assumption also contradicts the narrow framing feature shared by most investors (Barberis and Huang, 2009). Therefore, we do not impose any assumptions on the number of lagged observations that the investor will consider to predict future economic fundamentals. Instead, we use the data-driven method to determine the most appropriate choice of the lagged number of observations for China and the United States. We also consider the conditional homoscedasticity autoregressive model to match the first type of investor distorted beliefs in economic fundamentals. The model is given by

$$\tilde{z}_{t+1} - \mu_d = \sum_{j=1}^{P} \Gamma_j (\tilde{z}_{t+1-j} - \mu_d) + \tilde{\epsilon}_{t+1}, \quad \tilde{\epsilon}_t \sim i.i.d.N(0, \sigma_d^2), \tag{4.6}$$

where the value of the number of lagged observations p is determined by actual data. Several popular techniques are used for the choices of p, including the AIC and BIC methods. A detailed discussion on the choice of the value of p shall be presented discussed in the next section.

The subjective expectation deviates from the mathematical one; furthermore, the subjective expectation affects the calculation of the pricedividend ratios, which is given by

$$\tilde{\omega}_t(z_t) = \mathbf{E}^s[\beta(e^{(1-\gamma)\tilde{z}_{t+1}}(\tilde{\omega}_{t+1}(\tilde{z}_{t+1}) + 1)|\boldsymbol{I_t}]. \tag{4.7}$$

As DSGE models are becoming increasingly complex, we are moving away from a world where analytical or closed form solutions exist. To obtain a first impression of model implications, economists now rely heavily on numerical approximations. In this study, no closed form solution to Equation (6) can be found after incorporating the distorted beliefs. We consider the perturbation method with second-order Taylor expansion around some steady states. Approximation errors tend to occur when high moments are ignored. Cui and Hong (2015) propose a two-stage regression method in solving asset pricing models. This method can avoid the misspecification errors of the price-dividend ratios as sample size becomes large. They also establish a series GMM estimation method for Euler equations with recursive preferences (Cui and Hong, 2016). Given that majority of the related literature still considers the perturbation method as a powerful technique of approximating the price-dividend ratios, we adopt this solution technique to pin down the possible effects exhibited by different solving methods. In addition, given the simple time-discrete utility function this paper works with, we consider the perturbation method as an acceptable approximation procedure.

Here, we use $\tilde{R}_t$ to denote the gross return of the risky asset when incorporating distorted beliefs in the mean level of economic fundamentals, which is expressed as

$$\tilde{R}_{t+1}(z_{t+1}, z_t) = \frac{\tilde{\omega}_{t+1}(z_{t+1}) + 1}{\tilde{\omega}_t(z_t)} e^{z_{t+1}}. \tag{4.8}$$

We define the return of the risk-free asset $\tilde{R}_f$ as

$$\tilde{R}_{f,t} = \frac{1}{\beta} e^{\gamma(\Gamma z_t + (1-\Gamma)\mu_d) - \frac{\gamma^2 \delta_d^2}{2}}. \tag{4.9}$$

Model 2: Investor extrapolation biases in the volatility level of economic fundamentals

We consider another possible pattern of distorted beliefs of investors in economic fundamentals, namely, volatility clustering bias. Li and Hong (2011) discover a volatility clustering effect for the Chinese exchange rate market. When the Chinese exchange rate market experiences high volatility yesterday, it tends to suffer from another volatile market the next day. Li and Hong (2011) consider the autoregressive conditional heteroscedasticity model [AR(p)-ARCH(p)] to feature the volatility clustering effect. We borrow the same modeling method in capturing the volatility clustering type of distorted beliefs. The optimal number of lagged observations in each specific dataset is determined by using a number of information criteria, such as AIC and BIC. The objective dynamics of economic fundamentals is assumed to follow Equation 4.1. The volatility clustering type of investor distorted beliefs is specified as

$$\begin{cases} \tilde{z}_{t+1}-\mu & =\Gamma(\tilde{z}_t-\mu_d)+\varepsilon_{t+1} \\ \varepsilon_{t+1} & =\sqrt{h_t}u_{t+1} \\ h_t & =\alpha_0+\sum_{j=1}^{p}\alpha_j\varepsilon_j^2 \\ u_{t+1} & \sim \text{i.i.d.} \quad N(0,1) \end{cases} \tag{4.10}$$

where Γ is the persistence in mean, α_0 is the constant conditional volatility and α_j is the persistence in volatility. The price-dividend ratio under the volatility clustering bias can be written as:

$$\bar{\omega}_t=\bar{\omega}(z_t,\ldots,z_{t-2P+1})=\mathbf{E}^s[\beta(e^{(1-\gamma)\tilde{z}_{t+1}}(\bar{\omega}(\tilde{z}_{t+1},z_t,...,z_{t-2P})+1)|\boldsymbol{I_t}]. \tag{4.11}$$

The return on the risk-free asset at time t, which pays 1 unit of consumption at $t+1$, is:

$$\bar{R}_{f,t}=\frac{1}{\beta}e^{\gamma(\sum_{j=1}^{P}\Gamma z_{t+1-j}+(1-\sum_{j=1}^{P}\Gamma_j)\mu_d)-\frac{\gamma^2(\alpha_0+\sum_{j=1}^{P}\alpha_j(z_{t-j}-\mu_d-\sum_{k=1}^{P}\Gamma_k(z_{t-j-k}-\mu_d))^2)}{2}}. \tag{4.12}$$

Finally, the gross return on the risky asset at time $t+1$ is:

$$\bar{R}_{t+1}(z_{t+1},z_t,\ldots,z_{t-2P+1})=\frac{\bar{\omega}_{t+1}+1}{\bar{\omega}_t}e^{z_{t+1}}. \tag{4.13}$$

4.2 The GMM Estimation Method with Subjective Expectations

The GMM method, a well-known and widely applied estimation procedure in various economic studies, can only be used under the mathematical expectation. The full rationality assumption is implied to be a necessary condition to ensure correct estimation results. Consequently, models involving irrational expectations fail to perform estimations using GMM, and must rely on some calibration techniques. Given the specific research question raised in our paper, we must come up with enough calibration conditions to ensure just-identification requirement. This finding becomes relatively difficult, especially given the situation in which the number of lagged observations can be many and difficult to determine. Considering the relatively rich dataset we have to work with, a GMM-type estimation method can achieve better efficiency because of its ability to capture much detailed information compared with the calibration method. Thus, to obtain the optimal value of the number of lagged observation , and obtain efficient estimations of model parameters, we propose a modified GMM estimation method, which works with subjective expectations using some knowledge from measure theory. We first establish the following useful properties.

Proposition 1 (Change of measures). *Suppose E_t^s denotes some subjective expectations at time t, and E_t represents the mathematical expectation at time t. For some integrable and measurable functions $g_t(X_{t+1})$, such that a nonnegative integrable function exists, $h(\cdot)$ on $(-\infty, +\infty)$, which satisfies the following:*

$$E_t^s[g(X_{t+1})] = E_t[g(X_{t+1})h(X_{t+1})], \tag{4.14}$$

where $h(\cdot)$ is the Radon-Nikodym derivatives of subjective expectation with respect to the true mathematical expectations, and this is unique up to measure 0.

For the first type of distorted beliefs discussed in this paper, a corollary can be established for ease of estimations as follows:

Corollary 1. *Suppose z_t subjectively follows an $AR(p)$ type distorted process. Suppose its objective data generating process is $N(\mu_r, \sigma_r^2)$. Then, a nonnegative integrable function exists on $(-\infty, +\infty)$,*

$$h = \frac{\delta_r}{\delta_d} e^{\frac{(z_{t+1}-\mu_r)^2}{2\delta_r^2} - \frac{(z_{t+1}-\mu_d-\sum_{j=1}^{P}\Gamma_j(z_{t-j}-\mu_d)))^2}{2\delta_d^2}}, \tag{4.15}$$

such that for some integrable measurable functions, $g_t(z_{t+1})$, $E_t^s[g(X_{t+1})] = E_t[g(X_{t+1})h(X_{t+1})]$ holds.

If we consider the volatility clustering-type distorted beliefs discussed in the previous section, we can derive another useful corollary to assist our new GMM estimations.

Corollary 2. *Suppose z_t subjectively follows an $AR(p) - ARCH(p)$ type distorted process. Suppose its objective data generating process is $N(\mu_r, \sigma_r^2)$. Then, a nonnegative integrable function exists on $(-\infty, +\infty)$,*

$$h = \frac{\delta_r}{\delta_d} e^{\frac{(z_{t+1}-\mu_r)^2}{2\delta_r^2} - \frac{(z_{t+1}-\mu_d-\sum_{j=1}^{P}\Gamma(z_{t-j}-\mu_d)))^2}{2(\alpha_0+\sum_{k=1}^{P}\alpha_j(z_{t-k}-\mu_d-\sum_{j=1}^{P}(z_{t-j-k}-\mu_d)))^2}} \tag{4.16}$$

such that for some integrable measurable functions, $g_t(z_{t+1})$, $E_t^s[g(X_{t+1})] = E_t[g(X_{t+1})h(X_{t+1})]$ holds.

Next, we illustrate the process of implementing this new GMM estimation method in a subjective expectation using the concrete structure as studied in our paper. First, we consider the objective operating mechanism of economic fundamentals for China and the United States, respectively.

4.3 Data

China's stock market data are collected from the RESSET dataset. To facilitate fair comparison with the American stock market data, we consider the market-valued overall A stock returns as the risky asset. In terms of risk-free representative equity, we use a composite return in different time periods. From January to August 2002, we use 3-month deposit rate as the riskless return. From August 2002 to October 2006, we consider 3-month treasury bills as the risk-free asset. Since October 2006, we consider the return of Shanghai LIBOR as the risk-free return. Based on the equilibrium result that $D_t = C_t$, we choose consumption growth rate as an index for representing economic fundamentals in China. Given that consumption is a major component in overall GDP, and considering our direct contribution to investor utilities in most of the DSGE models, we consider consumption growth rate a convenient and a symbolic index for economic fundamentals. Inflation rate will be computed from the consumer price index. The American aggregate level data used in this paper are obtained from Robber Shillers work. We consider $S\&P$ 500 as the representative risky asset, and the return of the 6-month treasury bill as the return of risk-free asset. The Japan and UK's equity return data are retrieved from Kenneth French's database. Other related macro-data are downloaded from CEIC database.

The aforementioned returns are all nominal returns, which have not been adjusted for inflation rates. We obtain real returns using the method recommended by Cecchetti et al. (2000) and Campbell and Cochrane (1999). First, we derive inflation rates using the consumer price index for each country. Then, we regress inflation

Table 4.2: Parameter Values

Parameter	Notations	China	U.S.	Japan	U.K.
Mean of the consumption growth rate	μ_r	1.81%	0.3%	0.14%	0.23%
Standard deviation of the consumption growth rate	σ_r	2.07%	0.48%	1.07%	0.64%

Notes: All moments are in quarterly percentage.

rates on the past two observations of inflation rates, nominal returns of the risky and riskless assets to obtain the expected inflation rates. Finally, real returns for each asset can be accomplished by subtracting the expected inflation rates from the nominal returns.

For the rest of the parameters, including the time discount factor β, the relative risk aversion level γ, the persistence in the mean levels of economic fundamentals $\{\Gamma\}_{j=1}^{p}$, and the volatility clustering effect $\{\alpha_j\}_{j=1}^{q}$ with $p, q = 1, 2, \cdots$, we use a simulated method of moments (McFadden, 1989) to obtain consistent and unbiased estimations. Two types of moment conditions are employed in the estimation procedure. The first group of moment conditions includes differences between the moments generated by simulated data and the empirical data. Specifically, we let EP_{t+1} to denote the empirically observed equity premium at time t, $R_{t+1}$to denote the return of the risky asset, and $R_{f,t}$ for the return of the risk-free asset. Let $(\dot{E}P_{t+1}, \dot{R}_{t+1}, \dot{R}_{f,t})$ denote the corresponding values from the simulated data at each time period. When investors hold mean level extrapolation biases in economic fundamentals, $(\tilde{E}P_{t+1}, \tilde{R}_{t+1}, \tilde{R}_{f,t})$ specifies the distorted equity premium, return of the risky asset, and the risk-free asset. Meanwhile, $(\bar{E}P_{t+1}, \bar{R}_{t+1}, \bar{R}_{f,t})$

denotes the corresponding terms when the investor holds distorted beliefs in the mean and volatilities of economic fundamentals. Let $g_t^1(\theta, X)$ represent the first group of moment conditions, which measures the distance between the sample analogue of simulated data and the empirical data. For ease of notation, we denote $X \equiv (\dot{EP}_{t+1}, \dot{R}_{f,t}, EP_{t+1}, R_{t+1}, \frac{C_{t+1}}{C_t})$, which includes all the simulated data from different models and empirical observations given by

$$g^1(\theta, X) = R_{f,t} - \frac{1}{N}\sum_{t=1}^{N} \dot{R}_{f,t} \tag{4.17}$$

$$g^2(\theta, X) = (R_{f,t} - \frac{1}{T}\sum_{t=1}^{T} R_{f,t})^2 - \frac{1}{N}\sum_{t=1}^{N}(\dot{R}_{f,t} - \frac{1}{N}\sum_{t=1}^{N} \dot{R}_{f,t})^2 \tag{4.18}$$

$$g^3(\theta, X) = EP_{t+1} - \frac{1}{N-1}\sum_{t=1}^{N-1} \dot{EP}_{t+1} \tag{4.19}$$

$$g^4(\theta, X) = (EP_{t+1} - \frac{1}{T-1}\sum_{t=1}^{T-1} EP_{t+1})^2 - \frac{1}{N-1}\sum_{t=1}^{N-1}(\dot{EP}_{t+1} - \frac{1}{N-1}\sum_{t=1}^{N-1} \dot{EP}_{t+1})^2 \tag{4.20}$$

The second group of moment condition is the Euler equations from each specific model given by

$$\boldsymbol{E}_t^s[\beta \frac{c_{t+1}}{c_t}^{-\gamma} R_{t+1} - 1] = 0 \tag{4.21}$$

Due to the existence of extrapolation biases, the subjective expectation in Equation 4.23 is different from the mathematical expectation. As the original GMM estimation method becomes inapplicable here, we thus consider a modified GMM estimation method using Lemma 1, Corollary 1, and Corollary 2. The choice of instrumental variables is greatly simplified because of the property that the conditional mean is equal to zero. Therefore, all the information up to time t can be used as valid instrumental variables at each time t. Regardless of how large the number of lagged observations can be, this modified GMM estimation method can constantly obtain a just-or over-identified estimation. We let Z_t be a vector of instrumental variables with dimension large enough to ensure identification and specify the second group of moment conditions $g_t^2(\theta, X)$ given by

$$g_{1,t}^2(\theta, X) = [\beta(\frac{c_{t+1}}{c_t})^{-\gamma} R_{t+1} - 1] h(X_{t+1}) \bigotimes Z_t, \tag{4.22}$$

$$g_{2,t}^2(\theta, X) = [\beta(\frac{c_{t+1}}{c_t})^{-\gamma} R_{f,t} - 1] h(X_{t+1}) \bigotimes Z_t, \tag{4.23}$$

where $h_{t+1}(X_{t+1})$ is jointly determined by subjective expectations and the mathematical expectation. For the two distorted beliefs in economic fundamentals discussed in our paper, $h_{t+1}(X_{t+1})$ satisfies Corollary 1 and Corollary 2. Given that this modified GMM estimation method has no issue in choosing instrumental variables and achieving just-identified estimations, it serves as another contribution that our paper offers to the literature.

4.4 Estimation Results

We gradually add on extrapolation biases from complete rationality to the most appropriate level in order to visualize the effect of investor distorted beliefs on enhancing model performance. We first judge the model's explanation powers on equity

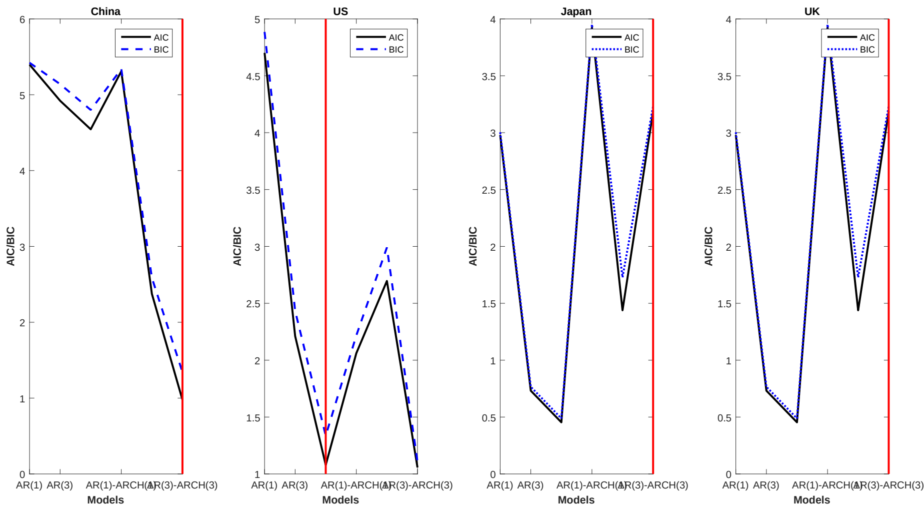


Figure 4.2: Model selections

premium puzzles for both China and the United States.

From Graph 4.2, both AIC and BIC information criteria indicate that the AR(3)-type distorted belief is the most suitable distortion pattern for the American stock market, whereas the AR(3)-ARCH(3) volatility clustering type is the best model for China's stock market. Notably, Chinese investors differ from American investors in many aspects. Our paper is able to capture these differences by exhibiting various types of distorted beliefs for investors from these two nations. American investors are highly confident regarding the overall stability of the country's economic background; thus, they are prone to react on single or short-term changes in economic fundamentals. Conversely, China's investors are more concerned with the uncertainty of what the macro economy looks like. Current Chinese reviving policies are more or less similar to the ones implemented in 2010 (Han and Hong, 2014), which have demonstrated short-term effectiveness. Furthermore, policies sharing this feature may be disabled in China, especially with the failure to meet the stability requirement from the general public. Another feature that makes China's stock market highly different from that of the United States is that over 90% of the investors in China are individual participants. They lack basic financial trainings and are more irrational in many aspects (Han and Hong, 2014). Even though the Chinese government has been attempting to improve the overall educational level of its stock market participants, it will take long to completely change current situations. Therefore, in the short-run, the Chinese government must face the irrationality shared by its general investors and implement policies that meet their fundamental concerns, such as the stability requirement, to help the stock market converge back to the real economy. From Table 3, investors' extrapolation biases in economic fundamentals clearly help explain equity returns. Seeing that the weight imposed by investors on historical observations is not exponentially decreasing is important. Thus, our paper convinces the argument in the beginning that assigning exponential decreasing weight to all past observations is inappropriate as Barberis et al. (2015) do in their paper. When investors from the United States are allowed to use three lagged historical data on economic fundamentals, the CAPM model with distorted beliefs has the best capacity to address all first two moments of risky and risk-free assets. The equity premium puzzle, which has confused the literature for decades, witnesses another reasonable explanation here. Table 3 Mean level extrapolation biases in explaining the equity premium puzzle

Table 4.3: Asset Returns with Extrapolation Biases

	Preferences		Beliefs									First and Second Moments				
	β	γ	μ_d	σ_d	Γ_1	Γ_2	Γ_3	α_0	α_1	α_2	α_3	μ_f	μ_{ep}	δ_f	δ_{ep}	$\rho_{ep,f}$
China	–	–	–	–	–	–	–	–	–	–	–	0.82	0.91	12.62	37.63	−0.27
Fully Rational	0.99	5.24	–	–	–	–	–	–	–	–	–	8.64	0	0.34	4.51	0
$AR(3) - ARCH(3)$	0.999	1.03	1.83%	–	0.0	0.217	0.0	3.73%	0.065	0.001	0.862	0.80	0.90	12.57	3	0
US	–	–	–	–	–	–	–	–	–	–	–	0.99	0.84	11.86	13.93	−0.20
Fully Rational	0.99	5.24	–	–	–	–	–	–	–	–	–	7.51	0	6.96	4.74	0
$AR(3)$	0.999	2.00	1.80%	0.01%	0.330	0.420	0.430	–	–	–	–	0.92	1.27	11.72	13.97	−0.53
Japan	–	–	–	–	–	–	–	–	–	–	–	1.12	0.83	5.14	17.35	0
Fully Rational	0.998	30	0.17%	1.05%	–	–	–	–	–	–	–	1.39	0.00	1.43	2.08	0
$AR(3)$	0.996	4.00	0.00%	1.6%	0.050	0.13	0.01	–	–	–	–	1.03	1.18	4.78	16.87	−0.1
UK	–	–	–	–	–	–	–	–	–	–	–	2.39	2.09	4.08	30.95	−0.19
Fully Rational	0.999	11.40	0.23%	0.64%	–	–	–	–	–	–	–	10.12	0.00	0.26	1.29	0.00
$AR(3)$	0.998	1.7	0.1%	0.2%	0.50	−0.15	0.70	–	–	–	–	2.55	1.90	4.04	31.03	−0.64

Note: Simulations are conducted at a quarterly level. All moments are in annual percentage. μ_d is the mean of distorted log consumption growth rate, with standard deviation measured by σ_d, both are in quarterly percentage. Columns named under preferences and extrapolation biases are estimated via the modified GMM estimation method discussed in the previous section. Columns named under the first two moments of equity returns represent the mean, volatility of equity premiums, and the risk-free asset under corresponding estimations of parameters. In addition, μ_f denotes the mean of the risk-free asset, with σ_f representing its standard error; μ_{ep} represents the mean of the equity premium, and σ_{ep} means its volatility; and features the correlation between the risk-free asset and the equity premium. All data used in this study are from 2002-2015.

Compared with the American representative investor, Chinese investors are not highly sensitive to changes in the mean level of economic fundamentals, as can be seen from Table 3. This finding helps address a puzzling phenomenon in China, wherein the stock market constantly deviates from its real economy; it further sheds some light on why too many positive government policies failed to revive China's stock market. China has enjoyed rapid development since the 1990s, during which the GDP, consumption growth, and the real estate market have all undergone unprecedented rates of expansion. Becoming accustomed to this fast-growing economy, investors tend to ignore marginal changes in how fast the economy is expanding. Therefore, changes in the mean level of economic fundamentals fail to create an influence on their stock market activities. For example, four trillion RMB was injected into China's real economy to revive seven targeting industries in 2010. Even though these targeting industries have shown significant improvement in production and benefits given to the real economy, the stock market still failed to show any positive attitude toward these government-supported industries at all (Han and Hong, 2014). Deeply confused by this phenomenon, the Chinese government has thought of ways to solve this problem. One possibility based on our simulation result is that the Chinese macroeconomic policies, such as the one implemented in 2010, generally last temporarily even if they are extremely strong. Chinese investors tend to not react to this type of instantaneous policies; thus, no matter how positive it seems to be, it has a huge chance of failure, at least in the stock market. This also provides a ground reason explaining why China's stock market deviates from the government's expectation and the status of the real economy. Next, we consider the other type of distorted beliefs, that is, the volatility clustering biases in economic fundamentals. We examine whether or not investors are concerned of the overall stable level of the macro economy. Table 4 reports the explanation powers on equity premiums under volatility type distortions. Chinese investors are extremely attuned towards changes in volatilities of economic fundamentals, whereas American investors have extremely weak reactions to it. When Chinese investors sense an unstable economic background, they tend to be extremely conservative regarding the stability situation in the future. Such attitude can further affect their trading strategies in the stock market. The introduction of volatility clustering-type distorted beliefs in economic fundamentals draws a better link between the simulated equity returns and the empirical returns. Specifically, the estimated mean of equity premium is approximately 2% higher than that from the empirical return, with 1% higher volatility. A dramatic difference between China and the United States is that the nominal risk-free return is not determined by the invisible handa finding that explains why our model is unable to capture the empirical return of the risk-free asset in China. However, some Chinese authority news media companies report that Chinese citizens consider financial products, such as mutual funds, as the actual risk-free assets in their perspective.

Two types of distorted beliefs in economic fundamentals with up to three lagged observations are studied in Tables 3 and 4, respectively. Questions may arise at this point: What is the best model for each nation, and what is the most suitable number of lagged observations used by each nation? In this study, we use AIC and BIC information criteria to answer these questions. Information criteria are computed under each model with different number of lagged observations, after which a model

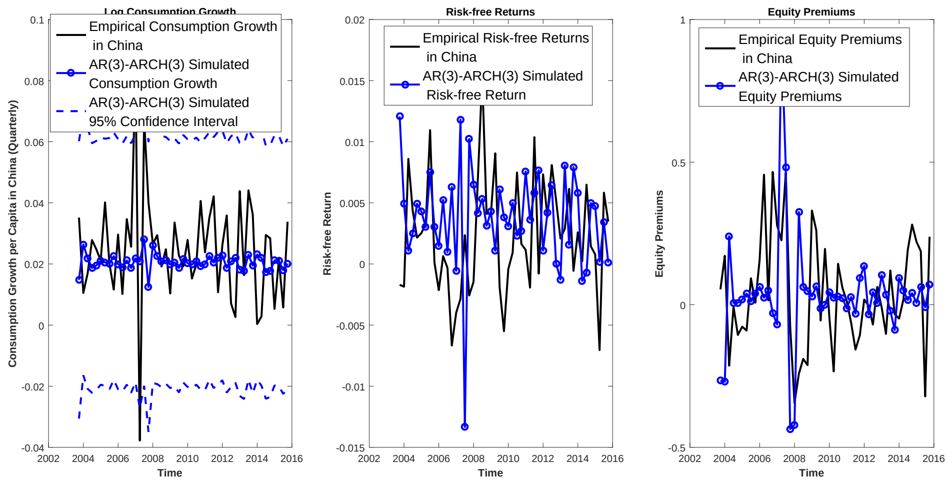


Figure 4.3: Equity returns from extrapolation biases in economic fundamentals and the empirical data in China

with a value of is uniquely chosen by choosing the pair that provides us the best fit between model complexity and estimation errors.

Using the most suitable model implied by real data, we prepare a time series comparison for equity returns between our models and the empirical data.

Generally, an investor's distorted beliefs on economic fundamentals definitely assist in drawing a deeper understanding of the stock markets in all four nations. Once pointed out, the equity premium puzzle has challenged the entire literature for long. Mehra and Prescott (1985) argue that the risk aversion level as high as 30, contradicts with empirical knowledge. While being able to recover all first two moments of the risk-free and risky asset, the risk aversion levels obtained under various distorted beliefs in the dynamics of the state variables are all much closer to

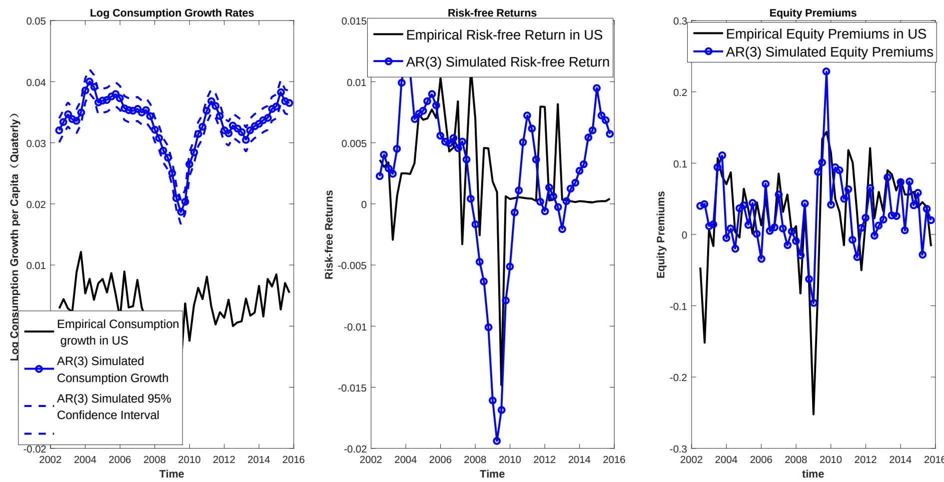


Figure 4.4: Equity returns from extrapolation biases in economic fundamentals and the empirical data in U.S.

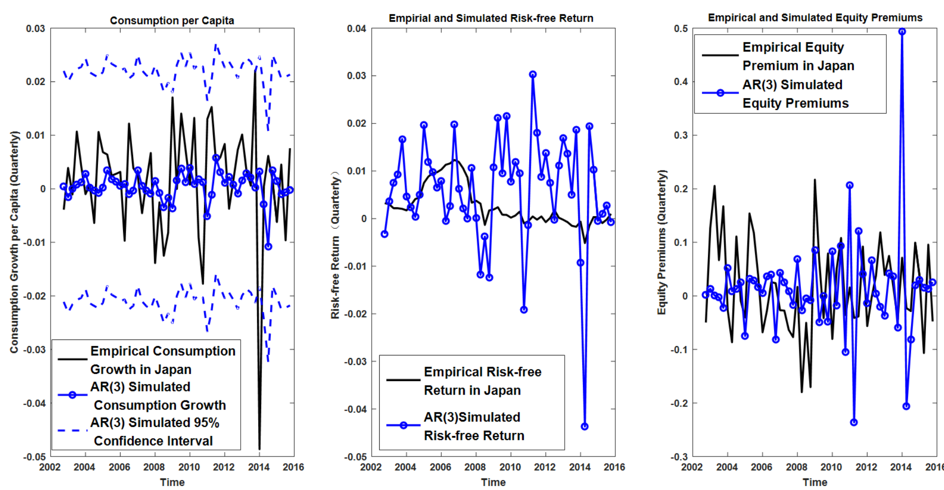


Figure 4.5: Equity returns from extrapolation biases in economic fundamentals and the empirical data in Japan

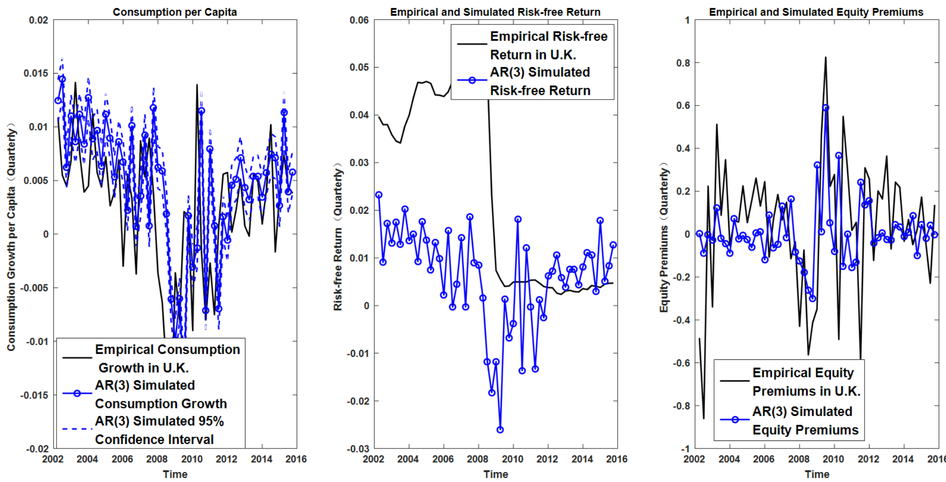


Figure 4.6: Equity returns from extrapolation biases in economic fundamentals and the empirical data in UK

how investors behave in reality. The parametrization implied in our paper is closer to reality; thus, the results launched in our paper has a better reliability for policy makers.

Next, we discuss our models' ability to predict future accumulative excess returns. Campbell and Shiller (1988) propose a loglinearized method in analysing the relationship between the accumulative excess returns and the dividend-price ratios. This method can be reduced to following regression model given by

$$r_{t+1} + r_{t+2} + ... + r_{t+j} = \alpha_j + \beta_j (D_t/P_t) + \epsilon_{j,t} \tag{4.24}$$

where, r_{t+1}denotes the return of the risky asset at time t, and β_j represents the marginal effect of the current period dividend-ratio on the next j periods accumula-

Table 4.4: Predictability of Accumulative Excess Returns

	China's Stock Market		U.S.'s Stock Market		Japan's Stock Market		U.K.'s Stock Market	
	$AR(3) - ARCH(3)$	Real Data	$AR(3)$	Real Data	$AR(3)$	Real Data	$AR(3)$	Real Data
β_1	9.326	4.512	1.13	1.734	1.484	2.456	4.152	4.113
β_2	17.553	11.281	2.342	3.764	2.626	4.984	6.643	7.384
β_3	22.696	19.479	3.490	5.810	3.824	8.850	10.331	11.853
β_4	27.751	24.019	4.603	7.847	4.979	12.477	11.856	15.940

Note: β_j are from regressions of j-year horizon cumulative stock returns on the lagged dividend-price ratio: $r_{t+1} + r_{t+2} + ... + r_{t+j} = \alpha_j + \beta_j(D_t/P_t) + \epsilon_{j,t}$.

tive excess returns. Based on the theory proposed by Campbell and Shiller (1988), we expect that the estimation of β_j increases with j. At this point, we run the above regression model using the respective simulated data from China and the United States, to test the performance of predictability of accumulative returns.

Both the mean level and volatility extrapolation biases for the United States and China perform well in predicting accumulative excess returns. However, the explanatory abilities of dividend-ratios on accumulative excess returns are relatively different. Considering that China's stock market is much younger, and participants are mainly composed of individual investors, other factors with a stronger influence on stock accumulative excess returns must exist. We are reminded that China requires customized policies to revive the market, instead of simply copying the successful experiences enjoyed by the American market.

4.5 Conclusions and Policy Implications

The traditional CAPM models have limited powers in addressing many well-documented facts, such as the equity premium puzzle and the predictability of the excess returns. By relaxing the Lucas Jr (1978) CAPM model and including investor extrapolation biases on economic fundamentals, our study makes the Lucas Jr (1978) model a special case and quantifies the differences in distortion types for investors in China and the United States. A distortion from historical fundamentals with endogenously determined weights on the past realizations can better explain the aggregate stock markets.

When the subjective expectation differs from the mathematical expectation, the traditional GMM procedure loses its power and fails to deliver correct estimations. Our paper overcomes this issue and proposes a new GMM estimation method that works with irrational expectations. Models involving irrational expectations can be calibrated using priori knowledge and can be estimated using as much information as we can from the empirical data. Our paper provides strong theoretical guidance for future work investigating irrationality.

Based on the data from 2002-2015, our paper explores the type of distorted beliefs that investors from China and the United States tend to hold. Furthermore, we find that American investors are insensitive to market volatility level, but are encouraged by improvements in the mean level of the economic fundamentals. In comparison, given that China is still under the developing stage, investors in the stock market care more regarding the overall stability situation of the macro economy. A major role played by the Chinese government is to design suitable economic policies to enhance the Chinese macro economy and the stock market (Hong, 2015). Many contemporaneous policies implemented by the Chinese government have failed to collect enough feedback from the stock market. Instead, based on the investor characteristics depicted in our paper, the Chinese government should consider those long-term and multi-stage policies. Only by understanding the type of investor irrationality can the Chinese government lead the stock market back to the trend of its real economy, and transform it into a more responsive and healthier financial market. In terms of improving the situation that Chinese median- and small-size industries' difficulties in financing, the Chinese government should consider lowering down the actual risk-free rate by stabilizing its policies as well. Considering the overwhelming number of imperfections exhibited by China's stock market, the government should consider enhancing the market using several strategies. First, information asymmetry must be resolved, and the government should attempt to ensure that the information is disclosed fairly to institutional traders and individual investors. Second, the Chinese government should continue promoting financial training activities to general investors. Third, many newly listed companies in China's stock market are actually poorly operated and are even back-door listed. Therefore, the Chinese government should reinforce its regulating policies to filter and block these listed companies away. Finally, Chinese monetary and financial policies should exhibit systematic, multi-stage, and sustainable properties. All these suggestions aim at enhancing investors' knowledge, reduce irrational trading behaviours, and enhance the ability that the government can guide the stock market.

APPENDIX A

APPENDIX OF CHAPTER 2

Note that Ψ_{Tp}, $\hat{\Psi}_{Tp}$ and Φ_{Tp} all depend on both T and p, but we suppress their subscript notations, using Ψ, $\hat{\Psi}$ and Φ respectively. ε has dimension equal T, and we also suppress its subscripts.

Proof of Theorem 3.3.2 According to Corollary 3.5 of Kress (1989), there exists a unique function $f^o \in G$ that solves $f - Af = \pi$ for each $\pi \in C(G)$. $\square$

Lemma A.0.1. *Since $f \in L^p$. Then there exists a continuous function $\bar{f}$ whose support lies in a bounded interval $[-A, A]$ so that*

$$||f - \bar{f}||_p < \epsilon.$$

Proof of Lemma B.0.1 This is immediate from the Stone-Weierstrass theorem. $\square$

Lemma A.0.2. *Suppose Assumption 3.3.2-3.3.5 hold. Then*

(a) $||\frac{\Phi'\Phi}{T} - E(\frac{\Phi'\Phi}{T})|| = O_p(\frac{p}{\sqrt{T}})$;

(b) $||\frac{\Psi'\Phi}{T} - E(\frac{\Psi'\Phi}{T})|| = O_p(\frac{p}{\sqrt{T}})$;

(c) with $\lambda(p)$ defined in Equation (2.19), under Assumption 3.3.6,

$\lambda_{\min}[\lambda(p)\frac{\Phi'\Phi}{T}] \rightarrow \lambda_{\min}[\lambda(p)E(\frac{\Phi'\Phi}{T})]$ *a.s.;*

(d) $\lambda_{\min}(\frac{\Phi'\Phi}{T}) > 0$ *a.s.;*

(e) $\lambda_{\max}(\frac{\Phi'\Phi}{T}) < \infty$ *a.s.*

Proof of Lemma B.0.2 Since the proof of (b) is analogous to (a), we only prove (a) here. We have

$$E[||\frac{\Phi'\Phi}{T}-E(\frac{\Phi'\Phi}{T})||^2]=\sum_{j=1}^{p}\sum_{i=1}^{p}E\{\sum_{l=0}^{T-1}\phi_i(X_l)\phi_j(X_l)/T-E[\sum_{l=0}^{T-1}\phi_i(X_l)\phi_j(X_l)/T]\}^2$$

$$\leq\sum_{j=1}^{p}\sum_{i=1}^{p}\{\frac{1}{T}\sup_{X_t\in X}var[\phi_i(X_t)\phi_j(X_t)]+2\sum_{0<k<m<T-1}cov[\phi_i(X_k)\phi_j(X_k),\phi_i(X_m)\phi_j(X_m)]/T^2\}$$

$$=A_1+A_2,\text{ say.}$$

Because $E||\phi_i(X_t)||_{4+\delta}<\Delta<\infty$ for some $\delta>0$ by assumption, we have $A_1\leq\frac{1}{T}\sum_{j=1}^{p}\sum_{i=1}^{p}E[\phi_i(X_t)\phi_j(X_t)]^2\leq\frac{1}{T}\sum_{j=1}^{p}\sum_{i=1}^{p}[E\big(\phi_i(X_t)\big)^4]^{\frac{1}{2}}[E\big(\phi_j(X_t)\big)^4]^{\frac{1}{2}}=O(\frac{p^2}{T})$. Given that $\{X_t\}$ is α mixing with coefficients $\alpha(n)$, by using the Davydov inequality and the condition on $E||\phi_i(X_t)||_{4+\delta}<\Delta<\infty$ for some $\delta>0$, we have

$$\begin{aligned}
A_2&=\frac{1}{T}\sum_{j=1}^{p}\sum_{i=1}^{p}\sum_{\tau=1}^{T-1}(1-\frac{\tau}{T})cov[\phi_i(X_t)\phi_j(X_t),\phi_i(X_{t+\tau})\phi_j(X_{t+\tau})]\\
&\leq\frac{2^{(4+2\delta)/(4+\delta)}(4+\delta)/\delta}{T}\sum_{j=1}^{p}\sum_{i=1}^{p}\sum_{\tau=1}^{T-1}|1-\frac{\tau}{T}|\alpha(\tau)^{\frac{\delta}{4+\delta}}\{E|\phi_i(X_t)\phi_j(X_t)|^{2+\delta/2}\}^{\frac{4}{4+\delta}}\\
&\leq\frac{c}{T}\sum_{j=1}^{p}\sum_{i=1}^{p}\sum_{\tau}^{T-1}\alpha(\tau)^{\frac{\delta}{4+\delta}}\{E|\phi_i(X_t)|^{4+\delta}E|\phi_j(X_t)|^{4+\delta}\}^{\frac{2}{4+\delta}}\\
&\leq\frac{Cp^2}{T}\sum_{\tau=1}^{\infty}\alpha(j)^{\frac{\delta}{4+\delta}}=O(\frac{p^2}{T}).
\end{aligned}$$

It follows that,

$$||\frac{\Phi'\Phi}{T}-E(\frac{\Phi'\Phi}{T})||=O_p(\frac{p}{\sqrt{T}})=o_p(1).$$

To obtain an almost sure convergence result for (c), we first establish a similar result under the convergence in probability. Similar to part (a), using the Markov, Cauchy-Schwarz and Holder inequalities, we have

$$
\begin{aligned}
&P[|\lambda_{\min}(\frac{\lambda(p)}{T}\sum_{t=0}^{T-1}\Phi_{qt}\Phi_{qt}') - \lambda_{\min}(\frac{\lambda(p)}{T}\sum_{t=0}^{T-1}E\Phi_{qt}\Phi_{qt}')| > \epsilon] \\
&\leq P\{\sum_{i=1}^{p}\sum_{j=1}^{p}|\frac{\lambda(p)}{T}\sum_{t=0}^{T-1}[\phi_i(X_t)\phi_j(X_t) - E\phi_i(X_t)\phi_j(X_t)]| > \epsilon\} \\
&\leq \frac{1}{\epsilon}\sum_{i=1}^{p}\sum_{j=1}^{p}E|\frac{\lambda(p)}{T}\sum_{t=0}^{T-1}[\phi_i(X_t)\phi_j(X_t) - E\phi_i(X_t)\phi_j(X_t)]| \\
&\leq \frac{1}{\epsilon}\sum_{i=1}^{p}\sum_{j=1}^{p}\big\{\frac{\lambda(p)^2}{T}\{\sup_{X\in X} var[\phi_i(X)\phi_j(X)] \\
&+ 2\sum_{0<k<m<T-1} cov[\phi_i(X_k)\phi_j(X_k), \phi_i(X_m)\phi_j(X_m)]/T\}\big\}^{\frac{1}{2}} \\
&= O_p(\frac{p^4\lambda(p)^2}{T})^{\frac{1}{2}}.
\end{aligned}
$$

Thus we have proved the convergence in probability for part (c). Conclusions under convergence in probability for (d) and (e) follow analogously. The proof of almost sure convergence follows Andrews (1991). □

Lemma A.0.3. *For p and T satisfying Assumption 3.3.4,*

(a) $\{\phi_{i,t}\psi_{j,\tau}\}$, $\{\Phi_{p,t}\varepsilon_{t+1}\}$, $\{\Psi_{qt}\varepsilon_{t+1}\}$ *and* $\{\Phi_{p,t}\Psi_{p,t}\}$ *are* α- *mixing sequences with coefficients* $\alpha(j)$*;*

(b) $E|\phi_{i,t}\psi_{j,t}|^{2r'} < \Delta' < \infty$ *for* $r' = r + \delta > 1$, $\forall t = 1, \cdots$ *and* $i, j = 1, \cdots, p$.

Proof of Lemma A.0.3 $\{\phi_{i,t}\}$ is a measurable function into **X**defined as a function of X_t. Because $\{X_t\}$ is assumed to be an α-mixing process of size $r/(r-1)$, $\{\phi_{i,t}\}$ is also an α-mixing process of size $r/(r-1)$ using the Thereom 3.49 of White (1996).

Similarly, $\{\psi_{i,t}\}$ is also an α-mixing process with size $r/(r-1)$. Immediately, from proposition 3.50 of White (1996), $\{\phi_{i,t}\phi_{j,\tau}\}$, $\{\phi_{i,t}\psi_{j,\tau}\}$, $\{\Phi_{p,t}\varepsilon_{t+1}\}$ and $\{\Psi_{q,t}\varepsilon_{t+1}\}$ are mixing sequences of size $r/(r-1)$.

For part (b), it immediately follows from the definition of $\psi_{i,t}$ and Minkowski's inequality that

$$
\begin{aligned}
&E|\phi_{i,t}\psi_{j,t}|^{2+\delta/2} = E|\phi_{i,t}\varphi_{j,t} - \phi_{i,t}\varphi_{j,t+1}m(X_{t+1})|^{2+\delta/2} \\
&\leq \{[E|\phi_{i,t}\varphi_{j,t}|^{2+\delta/2}]^{\frac{1}{2+\delta/2}} + [E|\phi_{i,t}\varphi_{j,t+1}m(X_{t+1})|^{2+\delta/2}]^{\frac{1}{2+\delta/2}}\}^{2+\delta/2} \\
&\leq \{[E|\phi_{i,t}^{4+\delta}|E|\varphi_{i,t}^{4+\delta}|]^{\frac{1}{4+\delta}} + [E|\phi_{i,t}|^{4+\delta}E|\varphi_{j,t}m(X_{t+1})|^{4+\delta}]^{\frac{1}{4+\delta}}\}^{2+\delta/2} < \Delta < \infty.
\end{aligned}
$$

$\square$

Lemma A.0.4. *Define* $G_p = E(\frac{\Psi'\Phi}{T})[E(\frac{\Phi'\Phi}{T})]^{-1}E(\frac{\Phi\Psi'}{T})$ *and* $G_{pT} = \frac{\hat{\Psi}'\hat{\Psi}}{T} = \frac{1}{T}\Psi'\Phi(\Phi'\Phi)^{-1}\Phi\Psi'$. *Suppose Assumptions 3.3.4-3.3.6 hold. Then*

(a) $||G_{pT} - G_p|| = O_p(\frac{\lambda(p)^2 p}{\sqrt{T}})$;

(b) $\lambda_{\max}(G_{pT}) = \lambda_{\max}(G_p) + O_p(\frac{\lambda(p)^2 p}{\sqrt{T}})$;

(c) $\lambda_{\min}(G_{pT}) \geq \frac{1}{2}\lambda_{\min}(G_p)$ *with probability approaching 1 as* $T \to \infty$.

Proof of Lemma B.0.3 For part (a), using the triangular inequality, we have

$$
\begin{aligned}
||G_{pT} - G_p|| &\leq ||(\frac{\Psi'\Phi}{T} - E\frac{\Psi'\Phi}{T})(\frac{\Phi'\Phi}{T})^{-1}\frac{\Phi\Psi'}{T}|| + ||E\frac{\Psi'\Phi}{T}[(\frac{\Phi'\Phi}{T})^{-1} - (E\frac{\Phi'\Phi}{T})^{-1}]\frac{\Phi'\Psi}{T}|| \\
&+ ||E\frac{\Psi'\Phi}{T}(E\frac{\Phi'\Phi}{T})^{-1}[\frac{\Phi'\Psi}{T} - E\frac{\Phi'\Psi}{T}]|| = A_3 + A_4 + A_5, \text{ say.}
\end{aligned}
$$

Using the results from Lemma A.0.3, we have

$$
A_3 \leq [\lambda_{\min}(\frac{\Phi'\Phi}{T})]^{-1}\lambda_{\max}(\frac{\Psi'\Phi}{T})||\frac{\Psi'\Phi}{T} - E\frac{\Psi'\Phi}{T}|| = O_p(\frac{\lambda(p)p}{\sqrt{T}}).
$$

Then, we show that $||(\frac{\Phi'\Phi}{T})^{-1} - (E\frac{\Phi'\Phi}{T})^{-1}|| \leq [\lambda_{\min}(\frac{\Phi'\Phi}{T})]^{-1}[\lambda_{\min}(E\frac{\Phi'\Phi}{T})]^{-1}||\frac{\Phi'\Phi}{T} - E\frac{\Phi'\Phi}{T}|| = O_p(\frac{\lambda(p)^2 p}{\sqrt{T}})$. Thus

$$A_4 \leq \lambda_{\max}(E\frac{\Psi'\Phi}{T})\lambda_{\max}(\frac{\Psi'\Phi}{T})||(\frac{\Phi'\Phi}{T})^{-1} - (E\frac{\Phi'\Phi}{T})^{-1}|| = O_p(\frac{\lambda(p)^2 p}{\sqrt{T}}).$$

Therefore, the last term

$$A_5 \leq \lambda_{\max}(E\frac{\Psi'\Phi}{T})(\lambda_{\min}E\frac{\Phi'\Phi}{T})^{-1}||\frac{\Phi'\Psi}{T} - E\frac{\Phi'\Psi}{T}|| = O_p(\frac{\lambda(p)p}{\sqrt{T}}).$$

It follows that $||G_{pT} - G_p|| = O_p(\frac{\lambda(p)^2 p}{\sqrt{T}})$.

Now we prove part (b):

$$\lambda_{\max}(G_{pT}) = \lambda_{\max}(G_p + G_{pT} - G_p) = \lambda_{\max}(G_p) + ||G_{pT} - G_p|| = \lambda_{\max}(G_p) + O_p(\frac{\lambda(p)^2 p}{\sqrt{T}}).$$

Next, we prove part (c). Similarly, we have

$$\lambda_{\min}(G_{pT}) \geq \lambda_{\min}(G_p) - ||G_{pT} - G_p|| \geq \lambda_{\min}(G_p) - O_p(\frac{\lambda(p)^2 p}{\sqrt{T}}) \geq \frac{1}{2}\lambda_{\min}(G_p).$$

□

Lemma A.0.5. *Suppose Assumptions 3.3.2-3.3.6 hold. Then there exists* $c_0 > 0$ *so that*

(a) $\lambda(p)\lambda_{\min}G_p \geq c_0 > 0$*;*

(b) $\lambda(p)\lambda_{\min}G_{pT} \geq \frac{c_0}{2} > 0\, a.s.$*;*

(c) $|||G_{pT}^{-1}(\frac{\Psi'\Phi}{T})(\frac{\Phi'\Phi}{T})^{-1} - G_p^{-1}E\frac{\Psi'\Phi}{T}(E\frac{\Phi'\Phi}{T})^{-1}|| = O_p(\frac{\lambda(p)^5 p}{\sqrt{T}})$

Proof of Lemma B.0.4 We first prove part (a). Denote a lead of matrix Φ as $\Phi_a = \sum_{t=1}^{T}\Phi_{p,t}\Phi'_{p,t}$ and a diagonal matrix $M = diag\{m_1, \cdots, m_T\}$. Recall the underlying structure of the asset pricing theory and our model construction. It is helpful to express $\Psi_{p,t} = \Phi_{p,t} - m_t\Phi_{p,t-1}$. Under Assumption 3.3.4, $E(m_t^2) = m < \infty$. Let $c, b \in R^p$ so that $c'(E\frac{\Psi'\Phi}{T})c = \lambda_{\min}(E\frac{\Psi'\Phi}{T})$, and $b'(E\frac{\Psi'\Phi}{T})b = \lambda_{\max}(E\frac{\Psi'\Phi}{T})$. Applying the Cauchy-Schwarz inequality and Lemma A.0.3, we have

$$\lambda_{\max}E(\frac{\Psi'\Phi}{T}) = b'E(\frac{\Phi'\Phi}{T})b - b'E(\sum_{t=0}^{T-1}\frac{\Phi_{t+1}\Phi'_t m_{t+1}}{T})b \le \lambda_{\max}E(\frac{\Phi'\Phi}{T}) + |b'E[\sup_{1\le t\le T}|m_t|\frac{\Phi'_a\Phi}{T}]b|$$
$$\le \lambda_{\max}E(\frac{\Phi'\Phi}{T}) + (E\sup_{1\le t\le T}m_t^2)^{\frac{1}{2}}b'[\frac{1}{T^2}E(\Phi'_a\Phi\Phi'\Phi_a)]^{\frac{1}{2}}b$$
$$= \lambda_{\max}E(\frac{\Phi'\Phi}{T}) + (E\sup_{1\le t\le T}m_t^2)^{\frac{1}{2}}b'[\frac{1}{T}E(\Phi'_a\Phi(\Phi'\Phi)^{-1}(\frac{\Phi'\Phi}{T})\Phi'\Phi_a)]^{\frac{1}{2}}b$$
$$\le \lambda_{\max}E(\frac{\Phi'\Phi}{T}) + \sqrt{\lambda_{\max}(\frac{\Phi'\Phi}{T})}(E\sup_{1\le t\le T}m_t^2)^{\frac{1}{2}}b'[\frac{1}{T}E(\Phi'_a\Phi(\Phi'\Phi)^{-1}\Phi'\Phi_a)]^{\frac{1}{2}}b$$
$$\le \lambda_{\max}E(\frac{\Phi'\Phi}{T}) + \sqrt{\lambda_{\max}(\frac{\Phi'\Phi}{T})}(E\sup_{1\le t\le T}m_t^2)^{\frac{1}{2}}\lambda_{\max}^{\frac{1}{2}}E(\frac{\Phi'_a\Phi_a}{T}) < \infty.$$

Therefore, $\lambda_{\max}E(\frac{\Psi'\Phi}{T}) = O_p(1)$. Because $\Phi(\Phi'\Phi)^{-1}\Phi'$ is an idempotent matrix, G_p is a square matrix and $\Psi'\Psi/T$ is invertible, we have

$$\lambda(p)\lambda_{\min}G_{p,T} \ge \lambda_{\min}[\Phi(\Phi'\Phi)^{-1}\Phi']\lambda(p)\lambda_{\min}[\Psi'\Psi/T] = 1$$

Using these facts, we can establish the following result that,

$$\lambda(p)\lambda_{\min}G_p \ge 1.$$

The almost sure convergence theorem in part (b) follows immediately by combining Lemma A.0.3 and B.0.3 together with Assumption 3.3.6. Finally, we prove part (c).

It is easy to show that

$$||G_p^{-1} - G_{pT}^{-1}|| = ||G_p^{-1}(G_p - G_{pT})G_{pT}^{-1}|| \leq [\lambda_{\min}(G_p)]^{-1}[\lambda_{\min}(G_{pT})]^{-1}||G_p - G_{pT}||$$
$$= O_p(\frac{\lambda^4(p)p}{\sqrt{T}}).$$

Plugging this result into the following inequality, we have

$$||G_{pT}^{-1}(\frac{\Psi'\Phi}{T})(\frac{\Phi'\Phi}{T})^{-1} - G_p^{-1}E\frac{\Psi'\Phi}{T}(E\frac{\Phi'\Phi}{T})^{-1}||$$
$$\leq ||(G_{pT}^{-1} - G_p^{-1})||\lambda_{\max}(\frac{\Psi'\Phi}{T})[\lambda_{\min}(\frac{\Phi'\Phi}{T})]^{-1} + ||\frac{\Psi'\Phi}{T} - (E\frac{\Psi'\Phi}{T})||[\lambda_{\min}(G_p)]^{-1}[\lambda_{\min}(E\frac{\Phi'\Phi}{T})]^{-1}$$
$$+ ||(\frac{\Phi'\Phi}{T})^{-1} - (E\frac{\Phi'\Phi}{T})^{-1}||[\lambda_{\min}(G_p)]^{-1}\lambda_{\max}(E\frac{\Psi'\Phi}{T})$$
$$= O_p(\frac{\lambda^5(p)p}{\sqrt{T}}).$$

□

Lemma A.0.6. *Suppose Assumptions 3.3.2-3.3.6 hold. Then*

(a) There exists a finite number $C > 0$ so that $E(\varepsilon\varepsilon') \leq CI_T$;

(b) $E[||\varphi^p(x)'\hat{\Psi}'\varepsilon/T||^2] = O_p(\frac{p}{T})$.

Proof of Lemma B.0.5 First, we prove part (a). Suppose an arbitrary vector $b = (b_1, b_2, \cdots, b_T)$ and a finite number $C > 0$ so that $C \geq c_\tau$, where $c_\tau = \Delta^{\frac{2}{4+\delta}} \sum_{\tau=0}^{\infty} \frac{2^{2-2/(4+\delta)}(4+\delta)}{2+\delta}\alpha(\tau)^{1-\frac{2}{4+\delta}}$ for some $\delta > 0$, and $\alpha(\tau)$ is the mixing coefficients. Then we have

$$b'(CI_T - E\varepsilon\varepsilon')b = C\sum_{t=1}^{T} b_t^2 - \sum_{t=1}^{T}\sum_{s=1}^{T} b_t b_s E(\varepsilon_t\varepsilon_s)$$
$$\geq C\sum_{t=1}^{T} b_t^2 - \frac{1}{2}\sum_{t=1}^{T}\sum_{s=1}^{T}(b_t^2 + b_s^2)E|\varepsilon_t\varepsilon_s| = C\sum_{t=1}^{T} b_t^2 - \sum_{t=1}^{T} b_t^2 \sum_{s=1}^{T} E|\varepsilon_t\varepsilon_s|$$
$$\geq C\sum_{t=1}^{T} b_t^2 - \sum_{\tau=0}^{\infty} \frac{2^{2-2/(4+\delta)}(4+\delta)}{2+\delta}\alpha(\tau)^{1-\frac{2}{4+\delta}}(E|\varepsilon_t|^{4+\delta})^{\frac{2}{4+\delta}}\sum_{t=1}^{T} b_t^2) \geq (C - c_\tau)\sum_{t=1}^{T} b_t^2.$$

Hence, $CI_T - E(\varepsilon\varepsilon')$ is positive semidefinite. Using the result from part (a) and Lemma B.0.3, we are can prove part (b) immediately:

$$E(||\varphi^p(x)'\hat{\Psi}'\varepsilon/T||^2) = \frac{1}{T}E\{tr[(\frac{\hat{\Psi}'\hat{\Psi}}{T})\varphi^p(x)'\varphi^p(x)\varepsilon'\varepsilon]\} \leq \lambda_{\max}(\frac{\hat{\Psi}'\hat{\Psi}}{T})\frac{1}{T}\lambda_{\max}(E\varepsilon\varepsilon')tr[\varphi^p(x)'\varphi^p(x)]$$

$$\leq O_p(1)\frac{p}{T}\lambda_{\max}(CI_T) = O_p(\frac{p}{T}).$$

□

Lemma A.0.7. *For each $x \in \mathbf{X}$, and $T \to \infty$, if $p > lnT$, and $(\frac{\xi_0(p)^2\lambda(p)lnT}{T})^{\frac{1}{2}} \to 0$,*

$$|\varphi^p(x)'(\frac{\hat{\Psi}'\hat{\Psi}}{T})^{-1}\hat{\Psi}\varepsilon/T| = O_p(\xi_o(p)\sqrt{\lambda(p)\frac{lnT}{T}}).$$

Proof of Lemma B.0.6 We shall show that for all $M > 0$ and $T \to \infty$,

$$P[(lnT)^{-\frac{1}{2}}[\lambda(p)]^{-\frac{1}{2}}T^{\frac{1}{2}}\xi_0(p)^{-1}|\varphi^p(x)'(\frac{\hat{\Psi}'\hat{\Psi}}{T})^{-1}\hat{\Psi}'\varepsilon/T| \leq M] \geq \frac{1}{2}.$$

Following Pollard (2012) and De Jong (2002), we have

$$P[(lnT)^{-\frac{1}{2}}[\lambda(p)]^{-\frac{1}{2}}T^{\frac{1}{2}}\xi_0(p)^{-1}|\varphi^p(x)'(\frac{\hat{\Psi}'\hat{\Psi}}{T})^{-1}\hat{\Psi}'\varepsilon/T| > M]$$

$$\leq 4P[(lnT)^{-\frac{1}{2}}[\lambda(p)]^{-\frac{1}{2}}T^{\frac{1}{2}}\xi_0(p)^{-1}|\sum_{t=1}^{T-1}\varphi^p(x)'(\frac{\hat{\Psi}'\hat{\Psi}}{T})^{-1}\hat{\Psi}_{p,t}\varepsilon_{t+1}\eta_{t+1}/T| > M/4] = 4P_1, \tag{A.1}$$

where $\eta = 1$ with probability $\frac{1}{2}$ and $\eta = -1$ with probability $\frac{1}{2}$. And we assume η_t is independent of ε and $\hat{\Psi}$. Apply the law of iterated expectations and Hoeffding's inequality, we have

$$P_1 = E(P_1|X,\varepsilon)$$
$$= E\{P[(lnT)^{-\frac{1}{2}}[\lambda(p)]^{-\frac{1}{2}}T^{\frac{1}{2}}\xi_0(p)^{-1}|\sum_{t=1}^{T-1}\varphi^p(x)'(\frac{\hat{\Psi}'\hat{\Psi}}{T})^{-1}\hat{\Psi}_{p,t}\varepsilon_{t+1}\eta_{t+1}/T| > M/4|X,\varepsilon]\}$$
$$\leq E\{2exp[-2(M/4)^2T(lnT)\lambda(p)\xi_0^2(p)/\sum_{t=1}^{T-1}[2\varphi^p(x)'(\hat{\Psi}'\hat{\Psi}/T)^{-1}\hat{\Psi}_{p,t}\varepsilon_{t+1}]^2\}$$
$$= P_2 \leq 1. \tag{A.2}$$

Then for all $c > 0$, by the Markov inequality, we have

$$P_2 \leq P[|\sum_{t=1}^{T-1}[2\varphi^p(x)'(\hat{\Psi}'\hat{\Psi}/T)^{-1}\hat{\Psi}_{p,t}\varepsilon_{t+1}]^2| > c^{-1}T\lambda(p)\xi_0^2(p)]$$
$$+ 2exp\{-2(M/4)^2 4^{-1}TlnT\lambda(p)/[c^{-1}T\lambda(p)\xi_0^2(p)]\}$$
$$\leq cT^{-1}\lambda(p)^{-1}\xi_0^{-2}(p)\varphi^p(x)'\lambda_{\max}E[(\hat{\Psi}'\hat{\Psi}/T)^{-1}\hat{\Psi}'\hat{\Psi}(\hat{\Psi}'\hat{\Psi}/T)^{-1}\varepsilon\varepsilon']\varphi^p(x) + O(-cM^2lnT/32)$$
$$\leq c + o(1). \tag{A.3}$$

Thus we obtain the result. □

Proof of theorem 3.3.4. In our 2SLS series regression procedure, an estimator of f_p is expressed as $\hat{f}_p(x) = \varphi^p(x)'\hat{\alpha}^p$. By the Minkowski inequality,

$$||\hat{f}_p - f|| = ||\hat{f}_p - f_p + f_p - f|| \leq ||\hat{f}_p - f_p|| + ||f_p - f|| \leq ||\hat{f}_p - f_p|| + O(p^{-s})$$

In the second step, we obtain $\hat{\alpha}^p = (\hat{\Psi}'\hat{\Psi})^{-1}\hat{\Psi}Y$. Consider the original time series nonlinear regression model, $\varepsilon_{t+1} = y_{t+1} - g_0(x_t, x_{t+1})$. Denote $G_0 = (g_0(x_0, x_1), \cdots, g_0(x_{T-1}, x_T))$. Under Assumption 3.3.2 that the variance of ε_{t+1} is finite, thus $E(\varepsilon\varepsilon') - \overline{\sigma}^2 I$ is positive semidefinite. We modify the proof of Theorem 1 of Newey (1994), and use the triangular inequality that

$$\int [\hat{f}_p(x) - f(x)]^2 dF(x) = \int [\hat{f}_p - f_p + f_p - f]^2 dF(x)$$
$$= \int [\varphi^p(x)'(\hat{\alpha}_p - \alpha_p) + \varphi^p(x)'\alpha_p - f]^2 dF(x) \leq ||\hat{\alpha}_p - \alpha_p||^2 + O(p^{-2s}).$$

Thus, we can focus on relevant properties of $||\hat{\alpha}_p - \alpha_p||$. It is immediately follows that,

$$||(\hat{\alpha}^p - \alpha^p)|| = ||(\hat{\Psi}'\hat{\Psi})^{-1}\hat{\Psi}'Y - (\hat{\Psi}'\hat{\Psi})^{-1}\hat{\Psi}'\hat{\Psi}\alpha^p||$$
$$\leq ||(\hat{\Psi}'\hat{\Psi})^{-1}\hat{\Psi}'Y - (\hat{\Psi}'\hat{\Psi})^{-1}\hat{\Psi}'G_0|| + ||(\hat{\Psi}'\hat{\Psi})^{-1}\hat{\Psi}'G_0 - (\hat{\Psi}'\hat{\Psi})^{-1}\hat{\Psi}'\hat{\Psi}\alpha^p||$$
$$\leq ||(\hat{\Psi}'\hat{\Psi})^{-1}\hat{\Psi}'(Y - G_0)|| + ||(\hat{\Psi}'\hat{\Psi})^{-1}\hat{\Psi}'(G_0 - \hat{\Psi}\alpha^p)||$$

Using the Cauch-Schwarz inequality, the property of an idempotent matrix, Lemma B.0.5 and Assumption 3.3.4, we obtain

$$E[||\varepsilon'\hat{\Psi}(\hat{\Psi}'\hat{\Psi}/T)^{-\frac{1}{2}}/T||^2] = trE[\hat{\Psi}(\hat{\Psi}'\hat{\Psi})^{-1}\hat{\Psi}\varepsilon\varepsilon']/T \leq tr\{E[\hat{\Psi}(\hat{\Psi}'\hat{\Psi})^{-1}\hat{\Psi}]\}^{\frac{1}{2}}\{E[\varepsilon\varepsilon']^2\}^{\frac{1}{2}}/T$$
$$\leq tr\{E[\hat{\Psi}(\hat{\Psi}'\hat{\Psi})^{-1}\hat{\Psi}]\}^{\frac{1}{2}}/T = O_p(\frac{p}{T}).$$

Therefore, it follows that

$$||(\hat{\Psi}'\hat{\Psi})^{-1}\hat{\Psi}'(Y - G_0)|| = ||(\hat{\Psi}\hat{\Psi}/T)^{-1}\hat{\Psi}'\varepsilon/T|| = |\varepsilon'\hat{\Psi}(\hat{\Psi}'\hat{\Psi}/T)^{-1}(\hat{\Psi}'\hat{\Psi}/T)^{-1}\hat{\Psi}'\varepsilon/T^2|^{\frac{1}{2}}$$
$$\leq \sqrt{\lambda(p)}||\varepsilon'\hat{\Psi}(\hat{\Psi}'\hat{\Psi}/T)^{-\frac{1}{2}}/T|| = O_p(\sqrt{\frac{\lambda(p)p}{T}}).$$

Recalling the construction of g_0 in Equation (2.15), we have

$$E|g_0(x_t, x_{t+1}) - \Psi'_{p,t}\alpha^p|^2 \leq CE|f(x_t) - f_p(x_{t+1})|^2 + CE|m(x_{t+1})[f(x_{t+1}) - f_p(x_{t+1})]|^2$$
$$= O(p^{-2s}).$$

By the Cauchy-Schwarz inequality and the fact that $\hat{\Psi}(\hat{\Psi}'\hat{\Psi})^{-1}\hat{\Psi}'$ is idempotent, we have

$$||(\hat{\Psi}'\hat{\Psi})^{-1}\hat{\Psi}'(G_0 - \Psi\alpha^p)|| = ||(G_0 - \Psi\alpha^p)'\hat{\Psi}(\hat{\Psi}'\hat{\Psi})^{-1}(\frac{\hat{\Psi}'\hat{\Psi}}{T})^{-1}\hat{\Psi}'(G_0 - \Psi\alpha^p)/T|^{\frac{1}{2}}$$
$$\leq O_P(\sqrt{\lambda(p)}p^{-s}).$$

Let $\hat{v} \equiv \Psi - \hat{\Psi}$, which is the estimated residual from the first stage OLS regression. The first order condition implies that $\Phi'\hat{v} = 0$. Given $\hat{\Psi} = \Phi(\Phi'\Phi)^{-1}\Phi'\Psi$, it is easy to show that $\hat{\Psi}'\hat{v} = \Psi'\Phi(\Phi'\Phi)^{-1}\Phi'\hat{v} = \mathbf{0}$. Then it immediately follows that

$$(\hat{\Psi}'\hat{\Psi})^{-1}\hat{\Psi}'(\Psi - \hat{\Psi})\alpha^p = (\hat{\Psi}'\hat{\Psi})^{-1}\Psi'\Phi(\Phi'\Phi)^{-1}\Phi'\hat{v} = \mathbf{0}.$$

Therefore, we conclude that $||\hat{\alpha}_p - \alpha_p|| = O_p[\sqrt{\lambda(p)}(\sqrt{\frac{p}{T}} + p^{-s})]$. It follows that,

$$\int [\hat{f}(x)_p - f(x)]^2 dF(x) = O_P[\lambda(p)(\frac{p}{T} + p^{-2s})].$$

Also, we can derive the rate of consistency of the 2SLS estimation under the Sobolev norm by applying the Cauchy-Schwarz inequality,

$$|\hat{f} - f|_d \leq |\varphi^p(\hat{\alpha}^p - \alpha^p)|_d + |f_p - f|_d \leq \xi_d(p)||\hat{\alpha}^p - \alpha^p|| + O_P(p^{-s})$$
$$= O_P[\xi_d(p)\sqrt{\lambda(p)}(\sqrt{\frac{p}{T}} + p^{-s})].$$

This rate is also derived in Newey (1997). De Jong (2002) establishes a sharper rate under the condition $p > lnT$. Following their reasoning, we can also obtain the same result. If $p > lnT$ and $\sqrt{\frac{\xi_d^2(p)\lambda(p)lnT}{T}} \to 0$, we have

$$|\hat{f}(x) - f(x)|_d \leq |\varphi^p(\hat{\alpha}^p - \alpha^p)|_d + |f_p - f|_d = |\varphi^p(x)'(\hat{\Psi}'\hat{\Psi})^{-1}\hat{\Psi}'\varepsilon|_d + O(p^{-s})$$
$$= O_p(\frac{\xi_d(p)\sqrt{\lambda(p)lnT}}{T}) + (1 + ||\Pi_p||_\infty)p^{-s}.$$

$\square$

Lemma A.0.8. *Define* $V_{pT}(x) = var[\frac{1}{\sqrt{T}}\varphi^p(x)'\Phi'\varepsilon] = \frac{1}{T}\varphi^p(x)'E(\Phi'\varepsilon\varepsilon'\Phi)\varphi^p(x)$. *Then as* $T \to \infty$,

$$\sqrt{\frac{1}{V_{pT}}}\varphi^p(x)'T^{-\frac{1}{2}}\Phi'\varepsilon \xrightarrow{d} N(0,1).$$

Proof of Lemma B.0.8 We prove the asymptotic normality by applying the martingale difference sequence central limit theorem by Brown (1971). First, we prove that for all fixed $x \in \mathbf{X}$, V_{pT} is well-defined. There exists $c \in R^p$ with $||c|| = 1$, we have

$$\begin{aligned} V_{pT}(x) &\equiv \varphi^p(x)'E[\Phi_{p,t}\Phi_{p,t}'\varepsilon_{t+1}^2]\varphi^p(x) \geq \varphi_{t+1}(x)'\lambda_{\min}E[\Phi_{p,t}\Phi_{p,t}'\varepsilon_{t+1}^2]\varphi_{t+1}(x) \\ &= \varphi_{t+1}^p(x)'\varphi_{t+1}^p(x)E[c'\Phi_{p,t}'\Phi_{p,t}\varepsilon_{t+1}^2 c] = O[\xi_0^2(p)]var[c'\Phi_{p,t}\varepsilon_{t+1}] = O[\xi_0^2(p)]. \end{aligned}$$

Because $E(\frac{1}{\sqrt{V_{pT}}}\varphi^p(x)'T^{-\frac{1}{2}}\Phi_{p,t}\varepsilon_{t+1}|I_t) = 0$, $\{\frac{1}{\sqrt{V_{pT}}}\varphi^p(x)'T^{-\frac{1}{2}}\Phi_{p,t}\varepsilon_{t+1}\}$ is a martingale difference sequence for all x and $t = 1, \cdots, T$. Second, we want to establish the Lindeberg condition given each x. By the Minkowski and triangular inequalities, we have

$$\begin{aligned} & V_{pT}(x)^{-1}T^{-1}\sum_{t=1}^{T} E\big\{[\varphi^p(x)'\Phi_{p,t}\varepsilon_{t+1}]^2 I\{[\varphi^p(x)'\Phi_{p,t}\varepsilon_{t+1}]^2 \geq \epsilon T V_{pT}\}\big\} \\ & \leq V_{pT}(x)^{-1}T^{-1}\sum_{t=1}^{T}(\epsilon T V_{pT})^{-\frac{\delta}{2}}E|\varphi^p(x)'\Phi_{p,t}\varepsilon_{t+1}]|^{2+\delta} \\ & \leq [V_{pT}(x)]^{-1-\frac{\delta}{2}}T^{-\frac{\delta}{2}}\{\sum_{i=1}^{p}\varphi_i(x)[E|\phi_{i,t}\varepsilon_{t+1}|^{2+\delta}]^{\frac{1}{2+\delta}}\}^{2+\delta} \\ & = o_p(1). \end{aligned}$$

The second condition that we need to verify that $\frac{1}{T}\sum_{t=0}^{T-1}\varphi^p(x)'\Phi_{p,t}\varepsilon_{t+1}^2\Phi_{p,t}'\varphi^p(x) -$

$V_{pT} = o_p(1)$. Given the fact that

$$E||\frac{\Phi'\varepsilon\varepsilon'\Phi}{T} - E\frac{\Phi'\varepsilon\varepsilon'\Phi}{T}|| = \sum_{i=1}^{p}\sum_{j=1}^{p} E[\frac{1}{T}\sum_{t=0}^{T-1}\phi_{i,t}\phi_{j,t}\varepsilon_{t+1}^2 - E\frac{1}{T}\sum_{t=0}^{T-1}\phi_{i,t}\phi_{j,t}\varepsilon_{t+1}^2]^2$$

$$\sum_{i=1}^{p}\sum_{j=1}^{p}[\frac{1}{T}E(\phi_{i,t}^2\phi_{j,t}^2\varepsilon_{t+1}^4) + \frac{2}{T^2}\sum_{1<k<m<T-1} cov(\phi_{i,t}\phi_{j,t}\varepsilon_{t+1}^2, \phi_{i,m}\phi_{j,m}\varepsilon_{m+1}^2)]$$

$$\leq \sum_{i=1}^{p}\sum_{j=1}^{p}[\frac{1}{T}\sqrt{E(\phi_{i,t}^4\phi_{j,t}^4)}\sqrt{E[\varepsilon_{t+1}^8|]}$$

$$+ \frac{2^{(4+2\delta)/(4+\delta)}(4+\delta)/\delta}{T}\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{\tau=1}^{T-1}\alpha(\tau)^{\frac{\delta}{4+\delta}}[E\phi_{i,t}^{4+\delta}\phi_{j,t}^{4+\delta}]^{\frac{2}{4+\delta}}[E|\varepsilon_{t+1}|^{8+2\delta}]^{\frac{2}{4+\delta}}$$

$$= O_p(\frac{p^2}{T}).$$

It immediately follows that

$$|\frac{1}{T}\sum_{t=0}^{T-1}\varphi^p(x)'\Phi_{p,t}\varepsilon_{t+1}^2\Phi_{p,t}'\varphi^p(x) - V_{pT}| = |tr\{\varphi^p(x)'(\frac{\Phi'\varepsilon\varepsilon'\Phi}{T} - E\frac{\Phi'\varepsilon\varepsilon'\Phi}{T})\varphi^p\}|$$

$$= |tr(\frac{\Phi'\varepsilon\varepsilon'\Phi}{T} - E\frac{\Phi'\varepsilon\varepsilon'\Phi}{T})\varphi^p(x)\varphi^p(x)'| \leq \lambda_{\max}|\frac{\Phi'\varepsilon\varepsilon'\Phi}{T} - E\frac{\Phi'\varepsilon\varepsilon'\Phi}{T}|p = O_p(\frac{p^3}{T}) = o_p(1).$$

It follows that $\frac{1}{\sqrt{V_{pT}}}\varphi^p(x)'\Phi'\varepsilon \xrightarrow{d} N(0,1)$ by Brown (1971). We observe the theorem.

□

Proof of theorem 3.3.5.

$D_{pT} = A_{pT}^{-2}(x) = \varphi^p(x)'(Q_T'P_T^{-1}Q_T)^{-1}Q_T'P_T^{-1}E(\Phi'\varepsilon\varepsilon'\Phi/T)P_T^{-1}Q_T(Q_T'P_T^{-1}Q_T)^{-1}\varphi^p(x)$.

As $T \to \infty$, based on Lemma A.0.3, we can drive a useful relationship between A_{pT} and V_{pT} that $O_p(1)V_{pT} \leq D_{pT} \leq O_p(\lambda(p)^4)V_{pT}$, which implies that $0 < O_p(\lambda(p)^{-2})V_{pT}^{-\frac{1}{2}} \leq A_{pT} \leq O_p(1)V_{pT}^{-\frac{1}{2}}$. Considering results from the 2SLS series regression, we have

$$\sqrt{T}\varphi^p(x)'(\hat{\alpha}^p - \alpha^p) = \sqrt{T}\varphi^p(x)'(\Psi'\Phi(\Phi'\Phi)^{-1}\Phi'\Psi)^{-1}\Psi'\Phi(\Phi'\Phi)^{-1}\Phi'\varepsilon.$$

Hence, by Lemmas B.0.4, B.0.8 and Assumption 2.3.4, we have

$$
\begin{aligned}
&|\sqrt{T}\varphi^p(x)'(\hat{\alpha}^p-\alpha^p)-\varphi^p(x)'(Q_T'P_T^{-1}Q_T)^{-1}Q_T'P_T^{-1}T^{-\frac{1}{2}}\Phi'\varepsilon| \\
&=|\varphi^p(x)'\{[\frac{\Psi'\Phi}{T}(\frac{\Phi'\Phi}{T})^{-1}\frac{\Phi'\Psi}{T}]^{-1}\frac{\Psi'\Phi}{T}(\frac{\Phi'\Phi}{T})^{-1}-(Q_T'P_T^{-1}Q_T)^{-1}Q_T'P_T^{-1}\}T^{-\frac{1}{2}}\Phi'\varepsilon| \\
&\leq \lambda_{\max}\{[\frac{\Psi'\Phi}{T}(\frac{\Phi'\Phi}{T})^{-1}\frac{\Phi'\Psi}{T}]^{-1}\frac{\Psi'\Phi}{T}(\frac{\Phi'\Phi}{T})^{-1}-(Q_T'P_T^{-1}Q_T)^{-1}Q_T'P_T^{-1}\}|T^{-\frac{1}{2}}\varphi^p(x)'\Phi'\varepsilon|V_{pT}^{-\frac{1}{2}}V_{pT}^{\frac{1}{2}} \\
&=O_p(\frac{\lambda^5(p)p}{T})O_p(1)O[\xi_0(p)]=o_p(1).
\end{aligned}
$$

It implies that $\sqrt{T}\varphi^p(x)'(\hat{\alpha}^p-\alpha^p)$ has the same limiting distribution as that of $\varphi^p(x)'(Q_T'P_T^{-1}Q_T)^{-1}Q_n'P_T^{-1}T^{-\frac{1}{2}}\Phi'\varepsilon$. It is sufficient to derive the limiting distribution of $\varphi^p(x)'(Q_T'P_T^{-1}Q_T)^{-1}Q_T'P_T^{-1}T^{-\frac{1}{2}}\Phi'\varepsilon$.

We apply Brown's (1971) CLT theorem for martingale difference sequences. It is easy to show that $E(A_{pT}\varphi^p(x)'(Q_T'P_T^{-1}Q_T)^{-1}Q_T'P_T^{-1}\Phi_{p,t}\varepsilon_{t+1}|I_t)=0$ for all $t=0,\cdots,T-1$. Define $\lambda_t=\varphi^p(x)'(Q_T'P_T^{-1}Q_T)^{-1}Q_T'P_T^{-1}e_i$ where $e_i\in R^p$ has the $i-$th element equal to 1 and 0 otherwise. By the Minkowski and Markov inequalities, the properties of trace, we have

$$
\begin{aligned}
&D_{pT}(x)^{-1}T^{-1}\sum_{t=1}^{T}E\big\{[\sum_{i=1}^{p}\lambda_{iT}\phi_{i,t}\varepsilon_{t+1}]^2I\{[\sum_{i=1}^{p}\lambda_{iT}\phi_{i,t}\varepsilon_{t+1}]^2\geq\epsilon TD_{pT}\}\big\} \\
&\leq D_{pT}^{-1}(x)T^{-1}\sum_{t=1}^{T}(\epsilon TD_{pT})^{-\frac{\delta}{2}}E|\sum_{i=1}^{p}\lambda_{iT}\phi_{i,t}\varepsilon_{t+1}|^{2+\delta} \\
&\leq D_{pT}^{-1-\frac{\delta}{2}}(x)T^{-\frac{\delta}{2}}\{\sum_{i=1}^{p}\lambda_{iT}[E|\phi_{i,t}\varepsilon_{t+1}|^{2+\delta}]^{\frac{1}{2+\delta}}\}^{2+\delta} \\
&\leq O(1)[D_{pT}]^{-1-\frac{\delta}{2}}T^{-\frac{\delta}{2}}|tr[\varphi^p(x)'(Q_T'P_T^{-1}Q_T)^{-1}Q_T'P_T^{-1}P_T^{-1}Q_T(Q_T'P_T^{-1}Q_T)^{-1}\varphi^p(x)]|^{1+\frac{\delta}{2}} \\
&\leq O(1)[D_{pT}]^{-1-\frac{\delta}{2}}T^{-\frac{\delta}{2}}[D_{pT}]^{1+\delta/2}\xi_0(p)^{-1-\frac{\delta}{2}}=o_p(1).
\end{aligned}
$$

It is straightforward to show that $var(A_{pT}\varphi^p(x)'(Q_T'P_T^{-1}Q_T)^{-1}Q_T'P_T^{-1}T^{-\frac{1}{2}}\Phi'\varepsilon)=1$. Thus using the same reasonings as in Lemma B.0.8, and given Assumption 2.3.5, we can show that

$$\frac{1}{T}\sum_{t=0}^{T-1}\varphi^p(x)'(Q_T'P_T^{-1}Q_T)^{-1}Q_T'P_T^{-1}\Phi_{pt}\varepsilon_{t+1}^2\Phi_{pt}'P_T^{-1}Q_T(Q_T'P_T^{-1}Q_T)^{-1}\varphi^p(x)-D_{pT}$$
$$=O_p(\frac{\lambda^4(p)p^3}{T})=o_p(1).$$

Thus we have proved that $A_{pT}\varphi^p(x)'(Q_T'P_T^{-1}Q_T)^{-1}Q_T'P_T^{-1}T^{-\frac{1}{2}}\Phi'\varepsilon \xrightarrow{d} N(0,1)$ as $T\to\infty$. Because we have proved that $A_{pT}\sqrt{T}\varphi^p(x)'(\hat{\alpha}^p-\alpha^p)$ has the same limiting distribution, it immediately follows that,

$$A_{pT}\sqrt{T}\varphi^p(x)'(\hat{\alpha}^p-\alpha^p)\xrightarrow{d} N(0,1),$$

which completes the proof of part (a).

For part (b) in Theorem 3.3.5, using the Slutsky theorem, it is sufficient to show that

$$A_{pT}\sqrt{T}[E\hat{f}_p(x)-f(x)]\to 0,\ \text{as}\, p,T\to\infty.$$

Define $f_p^r(x)=\sum_{j=p+1}^{\infty}\varphi_j(x)\alpha_j=f(x)-f_p(x)$ as the remainder part of the truncated series expansion f_p for f, where $f_p(x)=\sum_{j=1}^{p}\varphi_j(x)\alpha_j$. We further define

$$Q_{pT}\equiv A_{pT}\sqrt{T}[E\hat{f}_p(x)-f(x)]=A_{pT}\sqrt{T}f_p^r$$

Under Assumption 2.3.5 and Lemma B.0.8, we have

$$||Q_{pT}||\le\lambda_{\max}|A_{pT}\sqrt{T}|||f_p^r||=[\lambda_{\min}(D_{pT})]^{-\frac{1}{2}}\sqrt{T}O_p(p^{-s})$$
$$\le O_p(\sqrt{T}p^{-s})\frac{1}{\sqrt{\lambda_{\min}V_{pT}}}\le O_p(\sqrt{T}p^{-\frac{1}{2}-s})\to 0.$$

This implies that the approximation errors captured by Q_{pT} will be asymptotically negligible as $T\to\infty$. Therefore, we complete the proof that $A_{pT}\sqrt{T}[E\hat{p}(x)-f(x)]\to 0$,as $T\to\infty$. □

Proof of theorem 3.3.6 Recall the model setup that

$$\varepsilon_{t+1} = y_{t+1} - \sum_{i=1}^{\infty} \psi_{i,t}\alpha_i = y_{t+1} - \sum_{i=1}^{p} \psi_{i,t}\alpha_i - g_p^r.$$

We then have

$$\hat{\varepsilon}_{t+1} \equiv y_{t+1} - \sum_{i=1}^{p} \psi_{i,t}\hat{\alpha}_i = \varepsilon_{t+1} - \sum_{i=1}^{p} \psi_{i,t}(\hat{\alpha}_i - \alpha_i) + g_p^r,$$

where $g_p^r = o(1)$ by assumption. Define $S_{pT} \equiv E(\Phi'\varepsilon\varepsilon'\Phi/T) = \frac{1}{T}\sum_{t=1}^{T} E(\Phi_{p,t}\varepsilon_{t+1}^2\Phi'_{p,t})$ by properties of MDS. Let $\hat{S}_{pT} = \frac{1}{T}\sum_{t=1}^{T} \Phi_{p,t}\hat{\varepsilon}_{t+1}\hat{\varepsilon}_{t+1}\Phi'_{p,t}$ and $\bar{S}_{pT} = \frac{1}{T}\sum_{t=1}^{T} \Phi_{p,t}\varepsilon_{t+1}\varepsilon_{t+1}\Phi'_{p,t}$. Using triangular inequality, we have

$$||\hat{S}_{pT} - S_{pT}|| \leq ||\hat{S}_{pT} - \bar{S}_{pT}|| + ||\bar{S}_{pT} - S_{pT}||$$

$$= ||\frac{1}{T}\sum_{t=1}^{T} \Phi_{p,t}\hat{\varepsilon}_{t+1}^2\Phi'_{p,t} - \frac{1}{T}\sum_{t=1}^{T} \Phi_{p,t}\varepsilon_{t+1}^2\Phi'_{p,t}|| + ||\frac{1}{T}\sum_{t=1}^{T}(\Phi_{p,t}\varepsilon_{t+1}^2\Phi_{p,t} - E\Phi_{p,t}\varepsilon_{t+1}^2\Phi_{p,t})||$$

$$= B_1 + B_2, \text{ , say.}$$

The B_2 term has been shown to converge to zero in probability. Therefore, we only need to prove that $B_1 = o_p(1)$. By the triangular inequality, we have

$$B_1 = ||\frac{1}{T}\sum_{t=1}^{T}(\Phi_{p,t}\hat{\varepsilon}_{t+1}^2\Phi'_{p,t} - \Phi_{p,t}\varepsilon_{t+1}^2\Phi'_{p,t})|| \leq ||\frac{1}{T}\sum_{t=1}^{T}(\Phi_{p,t}\hat{\varepsilon}_{t+1} - \Phi_{p,t}\varepsilon_{t+1})(\Phi_{p,t}\hat{\varepsilon}_{t+1} - \Phi_{p,t}\varepsilon_{t+1})'||$$

$$+ 2||\frac{1}{T}\sum_{t=1}^{T}(\Phi_{p,t}\hat{\varepsilon}_{t+1} - \Phi_{p,t}\varepsilon_{t+1})\varepsilon_{t+1}\Phi'_{p,t}|| = B_{1,a} + B_{1,b}, \text{ say.}$$

To bound the $B_{1,a}$ term, we first observe that

$$\frac{1}{T}\sum_{t=1}^{T}(\hat{\varepsilon}_{t+1} - \varepsilon_{t+1})^2 \leq 2\frac{1}{T}\sum_{t=1}^{T}[\varphi^{p'}(\alpha^p - \hat{\alpha}^p)]^2 + 2\frac{1}{T}\sum_{t=1}^{T}||g_p^r||^2 = O_p\big(\lambda(p)(\frac{p}{T} + p^{-2s})\big) + o_p(1).$$

It follows from the Cauchy-Schwarz inequality that

$$E||B_{1,a}|| = \sum_{i=1}^{p}\sum_{j=1}^{p} E\{\sup_t \phi_{i,t}\phi_{j,t}\frac{1}{T}\sum_{t=1}^{T}(\hat{\varepsilon}_{t+1} - \varepsilon_{t+1})^2]\} \leq O_p(\lambda(p)(\frac{p}{T} + p^{-2s}))p^2 \sup_{i,t} E(\phi_{i,t})^2$$

$$\leq O_p\big(\lambda(p)(\frac{p}{T} + p^{-2s})p^2\big) = o_p(1).$$

For the $B_{1,b}$ term, Applying Jensen's inequality and the Cauchy-Schwarz inequality, we have

$$E||B_{1,b}|| = 2E||\frac{1}{T}\sum_{t=1}^{T}(\Phi_{p,t}\hat{\varepsilon}_{t+1} - \Phi_{p,t}\varepsilon_{t+1})\varepsilon_{t+1}\Phi'_{p,t}|| \leq 2E||\sup_t \Phi_{p,t}\Phi'_{p,t}\frac{1}{T}\sum_{t=1}^{T}(\hat{\varepsilon}_{t+1} - \varepsilon_{t+1})\varepsilon_{t+1}||$$
$$\leq 2E||\sup_t \Phi_{p,t}\Phi'_{p,t}||\sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{\varepsilon}_{t+1} - \varepsilon_{t+1})^2}\sqrt{\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{t+1}^2} = O_p\big(\sqrt{\lambda(p)}(\sqrt{\frac{p}{T}} + p^{-s})p^2\big) = o_p(1).$$

Note that if using the FFF series, the p^{-s} term which captures the approximation errors are arbitrarily small and can be dropped out. Therefore, under Assumption 2.3.5, we establish that $||S_{pT} - \hat{S}_{pT}|| = o_p(1)$. To complete the asymptotic normality, we need to construct the following two intermediate statistics,

$$\bar{D}_{pT}(x) = \varphi^p(x)'(Q'_T P_T^{-1} Q_T)^{-1} Q'_T P_T^{-1} \hat{S}_{pT} P_T^{-1} Q_T (Q'_T P_T^{-1} Q_T)^{-1} \varphi^p(x)$$

and $\bar{A}_{pT} = \bar{D}_{pT}^{-\frac{1}{2}}$. It follows immediately that $D_{pT}\bar{D}_{pT}^{-1} \xrightarrow{p} 1$ and $A_{pT}\bar{A}_{pT}^{-1} \xrightarrow{p} 1$.

Subsequently, to prove $A_{pT}^2\hat{D}_{pT} \xrightarrow{p} 1$, it suffices to show that $A_{pT}^2\hat{D}_{pT} - D_{pT}^{-1}\bar{D}_{pT} = o_p(1)$. Using the triangular inequality and previous results, we have

$$|A_{pT}^2\hat{D}_{pT} - D_{pT}^{-1}\bar{D}_{pT}| = A_{pT}^2||\hat{D}_{pT} - \bar{D}_{pT}||$$
$$= A_{pT}^2||\varphi^p(x)'G_{pT}^{-1}\frac{\Psi'\Phi}{T}(\frac{\Phi'\Phi}{T})^{-1}\hat{S}_{pT}(\frac{\Phi'\Phi}{T})^{-1}\frac{\Psi'\Phi}{T}G_{pT}^{-1}\varphi^p(x) - \varphi^p(x)'G_p^{-1}Q_T P_T^{-1}\hat{S}_{pT}P_T^{-1}Q_T G_p^{-1}\varphi^p(x)||$$
$$\leq A_{pT}^2||\varphi^p(x)'[G_{pT}^{-1}\frac{\Psi'\Phi}{T}(\frac{\Phi'\Phi}{T})^{-1} - G_p^{-1}Q_T P_T^{-1}]\hat{S}_{pT}(\frac{\Phi'\Phi}{T})^{-1}\frac{\Psi'\Phi}{T}G_{pT}^{-1}\varphi^p(x)||$$
$$+ A_{pT}^2||\varphi^p(x)'G_p^{-1}Q_T P_T^{-1}\hat{S}_{pT}[(\frac{\Phi'\Phi}{T})^{-1}\frac{\Psi'\Phi}{T}G_{pT}^{-1} - P_T^{-1}Q_T G_p^{-1}]\varphi^p(x)||$$
$$= O_p(1)A_{pT}^2||\varphi^p(x)||^2\lambda_{\max}(\hat{S}_{pT})||'G_{pT}^{-1}\frac{\Psi'\Phi}{T}(\frac{\Phi'\Phi}{T})^{-1} - G_p^{-1}Q_T P_T^{-1}||\lambda_{\max}((\frac{\Phi'\Phi}{T})^{-1})\lambda_{\max}(G_{pT}^{-1})$$
$$+ O_p(1)A_{pT}^2||\varphi^p(x)||^2\lambda_{\max}P_T^{-1}\lambda_{\max}(G_p^{-1})||(\frac{\Phi'\Phi}{T})^{-1}\frac{\Psi'\Phi}{T}G_{pT}^{-1} - P_T^{-1}Q_T G_p^{-1}||$$
$$= O_p(\frac{\lambda^7(p)p}{\sqrt{T}}) = o_p(1).$$

Finally, we complete the proof by showing that

$$|\hat{A}_{pT}^2 A_{pT}^{-2} - 1| = |\hat{D}_{pT}^{-1} D_{pT} - 1| = |\hat{D}_{pT}^{-1} D_{pT}||A_{pT}^2 \hat{D}_{pT} - 1| \xrightarrow{p} 0.$$

Thus, by the Slutsky theorem, we have

$$\sqrt{T}\hat{A}_{pT}[\hat{f}_p(x) - f(x)] = \sqrt{T}(\hat{A}_{pT} A_{pT}^{-1}) A_{pT}[\hat{f}_p(x) - f(x)] \xrightarrow{d} N(0, 1) \, \text{as} \, T \to \infty.$$

□

APPENDIX B

APPENDIX OF CHAPTER 3

For ease of derivations, we further specify the following notations for $t = 1, \cdots, T$,

$j = 1, \cdots, l$ and $h = \{f, g\}$. Denote

$m^l(U_t, h) = [m_1(U_t, h), \cdots, m_l(U_t, h)]'$;

$m_j(U_t, h):$ the jth moment condition;

$\hat{m}_{Tj}(h) = \frac{1}{T}\sum_{t=1}^{T} m_j(U_t, h):$ sample average of the jth moment condition;

$\hat{m}_T(h) = [\hat{m}_{T1}(h), \cdots, \hat{m}_{Tl}(h)]'$;

$\xi_j(U_t, h) = m_j(U_t, h) - Em_j(U_t, h)$;

$\xi_t(h) = [\xi_1(U_t, h), \cdots, \xi_l(U_t, h)]'$;

$\bar{g}_{Tj} = E[\frac{1}{T}\sum_{t=1}^{T} m_j(U_t, h)]'$;

$\zeta_{Tj}(h) = \sqrt{T}[\hat{m}_{Tj}(h) - E\hat{m}_{Tj}(h)] = \sqrt{T}\frac{1}{T}\sum_{t=1}^{T}\xi_j(U_t, h)$;

$\zeta_T(h) = [\zeta_{T1}(h), \cdots, \zeta_{Tl}(h)]'$;

$\hat{Q}_T(h, W) = \hat{m}_T(h)'W\hat{m}_T(h)$;

$\bar{Q}(h, W) = EQ_T(h, W)$;

$Q(h, W) = \lim_{T\to\infty}\hat{Q}_T(h, W)$;

$M_T(h, W) = Q_T(h, W) - \bar{Q}_T(h, W)$;

$T:$ is sample size;

$p = (p_{Tf}, p_{Th})':$ order of unknown functions;

$l = p_{Tf} + p_{Th}:$ number of moments.

Lemma B.0.1. *Under Assumptions 3.3.3, for each $k = 1, \cdots, l$, there exists some $c_k(U_t)$, for some $0 < \Delta < \infty$, such that $Ec_k(U_t) < \Delta < \infty$, and $|m_k(U_t, h_1) -$*

$m_k(U_t, h_2)| \leq c_k(U_t)||h_1 - h_2||$, *and* $\hat{m}_{Tk}(h_1) - \hat{m}_{Tk}(h_2) \lesssim ||h_1 - h_2||$.

Proof of Lemma B.0.1 By construction, $m_k(U_t, h) = Z_{r,t}e_t(f, g)$, where $e_{t,1}(f, g) = \beta^\theta(\frac{C_{t+1}}{C_t})^{\theta-\theta/\eta}e^{\theta g_{t+1}} - (e^{g(t)} - 1)^\theta$ and $e_{t,2}(f, h) = M_{t+1}(g)(\frac{D_{t+1}}{D_t})[e^{f_{t+1}} + 1] - e^{f_t}$. For any $h_1 = (f_1, g_1)$ and $h_2 = (f_2, g_2) \in \Theta_l$. Let $G_1 = e^{g_1}$, $G_2 = e^{g_2}$, $F_1 = e^{f_1}$, and $F_2 = e^{f_2}$. Given the assumption that $f_1, g_1, f_2, g_2 \in L^2_w$, we have $G_1, G_2, F_1, F_2 \in L^2_w$. Applying Taylor expansion on $f_{i,t} = \phi(x_t)'a^i$ and $g_{i,t} = \psi(x_t)'b^i$ with respect to a^i and b^i respectively, there exists $\tilde{a}$ and $\tilde{b}$, such that

$$|F_{2,t+1} - F_{1,t+1}| = |(a^2 - a^1)'\phi(x_{t+1})e^{\phi'_{t+1}\tilde{a}}| = |f_{2,t+1} - f_{1,t+1}|\tilde{F}_{t+1},$$

$$|G_{2,t+1} - G_{1,t+1}| = (b^2 - b^1)'\phi(x_{t+1})e^{\phi'_{t+1}\tilde{b}}| = |g_{2,t+1} - g_{1,t+1}|\tilde{G}_{t+1}.$$

Applying Taylor expansion and using the fact that $G_t = e^{g_t} = \frac{W_t}{C_t} > 1$ and $\theta < 0$, we have

$$|e_{t,1}(f_1, g_1) - e_{t,1}(f_2, g_2)| = |\beta^\theta(\frac{C_{t+1}}{C_t})^{\theta-\frac{\theta}{\eta}}[G^\theta_{1,t+1} - G^\theta_{2,t+1}] + [(G_{2,t} - 1)^\theta - (G_{1,t} - 1)^\theta]|$$
$$\leq |\theta|\beta^\theta(\frac{C_{t+1}}{C_t})^{\theta-\frac{\theta}{\eta}}|\tilde{G}_{t+1}||g_{1,t+1} - g_{2,t+1}| + |\theta||\bar{G}_t - 1|^{\theta-1}|\tilde{G}_t||g_{2,t} - g_{1,t}|$$

Consider instrumental variables $Z_{r,t} = [\varsigma_1(X_t), \cdots, \varsigma_r(X_t)]$, where $2r \geq p + q = l$. Using the triangular inequality, the Cauchy-Schwarz inequality, and Assumption 3.3.3, we have

$$|\varsigma_j(X_t)[e_{t,1}(f_1, g_1) - e_{t,1}(f_2, g_2)]| \leq C_{j,1}|g_{1,t+1} - g_{2,t+1}| + C_{j,2}|g_{1,t} - g_{2,t}|,$$

where,

$$E|C_{j,1}|^2 \equiv C \sup_{h_l\in\Theta_l} E[\varsigma^2_j(X_t)(\frac{C_{t+1}}{C_t})^{2\theta-2\frac{\theta}{\eta}}\tilde{G}^2_{t+1}] \leq \sup_{h_l\in\Theta_l}[E[\varsigma^8_j(X_t)]^{\frac{1}{4}}[E\tilde{G}^8_t]^{\frac{1}{4}}[E(\frac{C_{t+1}}{C_t})^{4\theta-4\frac{\theta}{\eta}}]^{\frac{1}{2}} < \infty,$$

$$E|C_{j,2}|^2 \equiv \sup_{h_l\in\Theta_l} E\{\varsigma^2_j(X_t)[|\tilde{G}_t - 1|^{2\theta-2}]|\tilde{G}^2_t|\} \leq E|\varsigma^8_j(X_t)|^{\frac{1}{4}}E|\tilde{G}^8_t|^{\frac{1}{4}}[E|\tilde{G}_t - 1|^{4(\theta-1)}]^{\frac{1}{2}} < \infty.$$

In the meanwhile,

$$
\begin{aligned}
&|e_{t,2}(f_1,g_1)-e_{t,2}(f_2,g_2)| = |\beta^\theta(\frac{C_{t+1}}{C_t})^{\theta-1-\frac{\theta}{\eta}}\frac{D_{t+1}}{D_t}(\frac{G_{1,t+1}}{G_{1,t}-1})^{\theta-1}(F_{1,t+1}+1)\\
&-F_{1,t}-\beta^\theta(\frac{C_{t+1}}{C_t})^{\theta-1-\frac{\theta}{\eta}}\frac{D_{t+1}}{D_t}(\frac{G_{2,t+1}}{G_{2,t}-1})^{\theta-1}(F_{2,t+1}+1)+F_{2,t}|\\
&\le |\beta^\theta(\frac{C_{t+1}}{C_t})^{\theta-1-\frac{\theta}{\eta}}\frac{D_{t+1}}{D_t}(\frac{G_{1,t+1}}{G_{1,t}-1})^{\theta-1}||(F_{1,t+1}-F_{2,t+1})|+|F_{1,t}-F_{2,t}|\\
&+|\beta^\theta(\frac{C_{t+1}}{C_t})^{\theta-1-\frac{\theta}{\eta}}\frac{D_{t+1}}{D_t}[(\frac{G_{1,t+1}}{G_{1,t}-1})^{\theta-1}-(\frac{G_{2,t+1}}{G_{2,t}-1})^{\theta-1}](F_{2,t+1}+1)=A_{0,1}+A_{0,2}
\end{aligned}
$$

Considering the relationship between F_t, G_t and f_t and g_t, we have

$$
\begin{aligned}
A_{0,1} &\equiv |\beta^\theta(\frac{C_{t+1}}{C_t})^{\theta-1-\frac{\theta}{\eta}}\frac{D_{t+1}}{D_t}(\frac{G_{1,t+1}}{G_{1,t}-1})^{\theta-1}||(F_{1,t+1}-F_{2,t+1})|+|F_{1,t}-F_{2,t}|\\
&\le |\beta^\theta(\frac{C_{t+1}}{C_t})^{\theta-1-\frac{\theta}{\eta}}\frac{D_{t+1}}{D_t}(\frac{G_{1,t+1}}{G_{1,t}-1})^{\theta-1}(1+\tilde{F}_{t+1})||f_{2,t+1}-f_{1,t+1}|+|\tilde{F}_t||f_{2,t}-f_{1,t}|
\end{aligned}
$$

The conclusion that $|\varsigma_j(X_t)A_{0,1} \le c_k|h_1-h_2|$ is satisfied as long as $|\varsigma_j(X_t)(\frac{C_{t+1}}{C_t})^{\theta-1-\frac{\theta}{\eta}}\frac{D_{t+1}}{D_t}(\frac{G_{1,t+1}}{G_{1,t}-1})^{\theta-1}(1+\tilde{F}_{t+1})|$ and $|\varsigma_j(X_t)\tilde{F}_t|$ are square integrable. By Assumption 3.3.3, the Markov inequality, and the law of iterated expectations, we have

$$
\begin{aligned}
&E[\varsigma_j(X_t)\beta^\theta(\frac{C_{t+1}}{C_t})^{\theta-1-\frac{\theta}{\eta}}\frac{D_{t+1}}{D_t}(\frac{G_{1,t+1}}{G_{1,t}-1})^{\theta-1}(1+\tilde{F}_{t+1})]\\
&=E\{\varsigma_j(X_t)E_t[\beta^\theta(\frac{C_{t+1}}{C_t})^{\theta-1-\frac{\theta}{\eta}}\frac{D_{t+1}}{D_t}(\frac{G_{1,t+1}}{G_{1,t}-1})^{\theta-1}(1+\tilde{F}_{t+1})]\}\\
&=E[\varsigma_j(X_t)\tilde{F}_t]\le[E\varsigma_j^2(X_t)]^{\frac{1}{2}}[E\tilde{F}_t^2]^{\frac{1}{2}}<\infty.
\end{aligned}
$$

Analogously, using the triangular inequality, we have

$$
\begin{aligned}
A_{0,2} &\equiv |\beta^{\theta}(\frac{C_{t+1}}{C_t})^{\theta-1-\frac{\theta}{\eta}}\frac{D_{t+1}}{D_t}[(\frac{G_{1,t+1}}{G_{1,t}-1})^{\theta-1}-(\frac{G_{2,t+1}}{G_{2,t}-1})^{\theta-1}](F_{2,t+1}+1) \\
&= |\beta^{\theta}(\frac{C_{t+1}}{C_t})^{\theta-1-\frac{\theta}{\eta}}\frac{D_{t+1}}{D_t}[\frac{G_{1,t+1}}{G_{1,t}-1})^{\theta-1}-(\frac{G_{1,t+1}}{G_{2,t}-1})^{\theta-1}+(\frac{G_{1,t+1}}{G_{2,t}-1})^{\theta-1}-(\frac{G_{2,t+1}}{G_{2,t}-1})^{\theta-1}](F_{2,t+1}+1) \\
&\leq |\beta^{\theta}(\frac{C_{t+1}}{C_t})^{\theta-1-\frac{\theta}{\eta}}\frac{D_{t+1}}{D_t}[(\frac{G_{1,t+1}}{G_{1,t}-1})^{\theta-1}-\frac{G_{1,t+1}}{G_{2,t}-1})^{\theta-1}](F_{2,t+1}+1) \\
&+ |\beta^{\theta}(\frac{C_{t+1}}{C_t})^{\theta-1-\frac{\theta}{\eta}}\frac{D_{t+1}}{D_t}[(\frac{G_{1,t+1}}{G_{2,t}-1})^{\theta-1}-(\frac{G_{2,t+1}}{G_{2,t}-1})^{\theta-1}](F_{2,t+1}+1) \\
&\leq |(1-\theta)\beta^{\theta}(\frac{C_{t+1}}{C_t})^{\theta-1-\frac{\theta}{\eta}}\frac{D_{t+1}}{D_t}G_{1,t+1}^{\theta-1}(G_{1,t}-1)^{-\theta}G_{1,t}(F_{2,t+1}+1)|g_{2,t}-g_{1,t}| \\
&+ |\beta^{\theta}(\frac{C_{t+1}}{C_t})^{\theta-1-\frac{\theta}{\eta}}\frac{D_{t+1}}{D_t}(G_{2,t}-1)^{1-\theta}[G_{1,t+1}^{\theta-1}-G_{2,t+1}^{\theta-1}](F_{2,t+1}+1) \\
&\leq |(1-\theta)\beta^{\theta}(\frac{C_{t+1}}{C_t})^{\theta-1-\frac{\theta}{\eta}}\frac{D_{t+1}}{D_t}G_{1,t+1}^{\theta-1}(G_{1,t}-1)^{1-\theta}(F_{2,t+1}+1)[|g_{2,t}-g_{1,t}|+o(g_{2,t}-g_{1,t})^2] \\
&+ |(1-\theta)\beta^{\theta}(\frac{C_{t+1}}{C_t})^{\theta-1-\frac{\theta}{\eta}}\frac{D_{t+1}}{D_t}(G_{2,t}-1)^{1-\theta}G_{2,t+1}^{\theta-1}(F_{2,t+1}+1)[|g_{2,t+1}-g_{1,t+1}|+o(g_{2,t+1}-g_{1,t+1})^2]
\end{aligned}
$$

It follows immediately that,

$$
\begin{aligned}
&E|\varsigma_j(X_t)\beta^{\theta}(\frac{C_{t+1}}{C_t})^{\theta-1-\frac{\theta}{\eta}}\frac{D_{t+1}}{D_t}G_{1,t+1}^{\theta-1}(G_{1,t}-1)^{1-\theta}(F_{2,t+1}+1)| \\
&= E[\varsigma_j(X_t)E_t[\beta^{\theta}(\frac{C_{t+1}}{C_t})^{\theta-1-\frac{\theta}{\eta}}\frac{D_{t+1}}{D_t}G_{1,t+1}^{\theta-1}(G_{1,t}-1)^{1-\theta}(F_{2,t+1}+1)]] < \infty
\end{aligned}
$$

And similarly,

$$
E[\varsigma_j(X_t)[\beta^{\theta}(\frac{C_{t+1}}{C_t})^{\theta-1-\frac{\theta}{\eta}}\frac{D_{t+1}}{D_t}(G_{2,t}-1)^{\theta-1}G_{2,t+1}^{\theta-1}(F_{2,t+1}+1)]] = E[\varsigma_j(X_t)F_{2,t}] < \infty
$$

Therefore, we have proved that there exists a $c_j(X_t) \in L_w^1$, for all $j = 1\cdots, l$, such that

$$
|m_j(U_t,h_1) - m_j(U_t,h_2)| \leq c_j(X_t)|h_1 - h_2|.
$$

In addition, it is easy to show that under Assumption 3.3.3, $c_j(X_t)$ is square integrable, namely $c_j(X_t) \in L_w^2$. We prove the last part of this lemma using triangular inequality.

$$|\hat{m}_{Tj}(h_1) - \hat{m}_{Tj}(h_2)| = |\frac{1}{T}\sum_{t=1}^{T} m_j(U_t, h_1) - m_j(U_t, h_2)|$$
$$\leq \frac{1}{T}\sum_{t=1}^{T} |m_j(U_t, h_1) - m_j(U_t, h_2)| \leq \sup_t c_j(X_t)|h_1 - h_2|$$

Thus, we complete the proof. □

Lemma B.0.2. *Under Assumptions 3.3.2 and 3.3.3,* $\frac{1}{\sqrt{T}}(\zeta_T'\zeta_T - E\zeta_T'\zeta_T) \to^p 0$ *uniformly in* $h \in \Theta$ *for* $l \to \infty$ *as* $T \to \infty$.

Proof of Lemma B.0.2 Note that

$$\frac{1}{l}(\zeta_T'\zeta_T - E\zeta_T'\zeta_T) = \frac{1}{l}\sum_{j=1}^{l}[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi_j(U_t, h)]^2 - \frac{1}{l}\sum_{j=1}^{l}E[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi_j(U_t, h)]^2$$
$$= \frac{1}{l}\sum_{j=1}^{l}[\frac{1}{T}\sum_{t=1}^{T}\left(\xi_j(U_t, h)^2 - E\xi_j(U_t, h)^2\right)] + \frac{1}{l}\sum_{j=1}^{l}\frac{1}{T}\sum_{s\neq t}[\xi_j(U_s, h)\xi_j(U_t, h) - E\xi_j(U_s, h)\xi_j(U_t, h)]$$
$$= \frac{1}{T}\sum_{t=1}^{T}[\frac{1}{l}\xi_t(h)'\xi_t(h) - E\frac{1}{l}\xi_t(h)'\xi_t(h)] + \frac{1}{T}\sum_{s\neq t}[\frac{1}{l}\xi_s'(h)\xi_t(h) - \frac{1}{l}E\xi_s'(h)\xi_t(h)]$$
$$= L_1^{(1)} + L_1^{(2)}.$$

We prove the statement by showing $\frac{l}{\sqrt{T}}L_1^{(i)} \to^p 0$ for $i = 1, 2$ respectively.

First, it is straightforward that $EL_1^{(1)} = 0$. What's more, with Assumption 3.3.2-

3.3.3,the Daydov's inequality and the Minkowksi's inequality, we have

$$var(L_1^{(1)}) = \frac{1}{T^2}\sum_{t=1}^{T} var[\frac{1}{l}\sum_{j=1}^{l}\xi_j^2(U_t,h)] + \frac{1}{T^2}\sum_{s\neq t} cov[\frac{1}{l}\sum_{j=1}^{l}\xi_j^2(U_s,h), \frac{1}{l}\sum_{j=1}^{l}\xi_j^2(U_t,h)]$$

$$\leq \frac{1}{T} var(\frac{1}{l}\sum_{j=1}^{l}\xi_j^2(U_t,h))$$

$$+ \frac{1}{T}\sum_{\tau=1}^{T-1}(1-\frac{\tau}{T})\beta(\tau)^{\frac{\eta}{4+\eta}}[E\big(\frac{1}{l}\sum_{k=1}^{l}\xi_k^2(U_s,h)\big)^{2+\eta/2} E\big(\frac{1}{l}\sum_{j=1}^{l}\xi_k^2(U_{s+\tau},h)\big)^{2+\eta/2}]^{\frac{2}{4+\eta}}$$

$$\leq \frac{1}{Tl}\sup_{1\leq j\leq l} E[\xi_j^4(U_t,h) + 2\sum_{\tau=1}^{\infty}\beta(\tau)^{\frac{\eta}{4+\eta}}(E\xi_j^{4+\eta})^{\frac{4}{4+\eta}}]$$

$$+ \frac{1}{T}\sup_{1\leq j\leq l}\sum_{\tau=1}^{\infty}\beta(\tau)^{\frac{\eta}{4+\eta}} E[\xi_j^{4+\eta}(U_t,h)]^{\frac{4}{4+\eta}}$$

$$= \mathbb{O}_p(\frac{1}{T}) = o_p(1).$$

Thus, by Chebyshev's inequality, we have $L_1^{(1)} \to^p 0$.

Second, using the law of iterated expectations, we can show that $EL_1^{(2)} = 0$. Given Assumption 3.3.2-3.3.3, the Daydov's inequality and the Minkowksi's inequality, we have

$$\frac{l^2}{T} var(L_1^{(2)}) = \frac{l^2}{T}\frac{4}{l^2T^2} E[\sum_{s<t}\xi_s(h)'\xi_t(h)\sum_{m<n}\xi_m(h)'\xi_n(h)]$$

$$= \frac{l^2}{T}\frac{4}{T^2}\sum_{s<t}\sum_{m<n}\sup_{j,s,t} E[\xi_j(U_s,h)\xi_j(U_t,h)]^2 = O_p(\frac{l^2}{T}) = o_p(1).$$

□

Lemma B.0.3. *Consider the $l \times l$ optimal weighting matrix $\Omega(h) = E[m(U_t,h)m(U_t,h)']$. Under Assumptions 3.3.2-3.3.3, there exists some $0 < \Delta < \infty$ so that for any nonstochastic vectors $a, b \in \mathbb{R}^l$, $|a'\Omega(h_1)b - a'\Omega(h_2)b| \leq C||a||||b||||h_1 - h_2||$.*

Proof of Lemma B.0.3 Using the Cauchy-Schwarz inequality, the c_r inequality, Lemma B.0.1 and Lemma B.0.2, we have

$$|a'\Omega(h_1)b - a'\Omega(h_2)b| = E|a'b[m^l(U_t,h_1)'m(U_t,h_1) - m^l(U_t,h_2)'m(U_t,h_2)]|$$

$$\leq \sup_{h\in\Theta} E|a'bm^l(U_t,h)'C(U_t)|||h_1-h_2|| \leq \sup_{h\in\Theta}[E|a'm^l(U_t,h)|^2]^{\frac{1}{2}}[E|b'c(U_t)|^2]^{\frac{1}{2}}$$

$$\leq c_r^2[\sum_{j=1}^{l} a_j^2 E|m_j(U_t,h)|^2]^{\frac{1}{2}}[\sum_{j=1}^{l} b_j^2 E|c_j(U_t)|^2]^{\frac{1}{2}}||h_1-h_2||$$

$$\leq C||a||||b||||h_1-h_2||.$$

□

Lemma B.0.4. *Under Assumption 3.3.3, uniformly in all $h \in \Theta_l$,*

$$\sup_{h\in\Theta_l} ||\frac{1}{T}\sum_{t=1}^{T} m^l(U_t,h)m^l(U_t,h)' - E[m^l(U_t,h)m^l(U_t,h)']|| \xrightarrow{p} 0$$

Proof of Lemma B.0.4 Under Assumption 3.3.3 that $E|m_j(U_t,h)|^{4+\eta} \leq \Delta < \infty$ for some $\eta > 0$, we have

$$E||\hat{W}(h) - W(h)||^2 = E|\sum_{i=1}^{l}\sum_{j=1}^{l}[\frac{1}{T}\sum_{t=1}^{T}\hat{W}_{t,(i,j)}(h) - EW_{i,j}(h)]^2|$$

$$\leq \sum_{i=1}^{l}\sum_{j=1}^{l}\frac{1}{T}var(W_{t,(i,j)}(h)) + \sum_{i=1}^{l}\sum_{j=1}^{l}\frac{2}{T^2}\sum_{s\neq t} cov[W_{t,(i,j)}(h), W_{s,(i,j)}(h)]$$

$$\leq O(\frac{l^2}{T}) + \frac{l^2}{T}\sum_{\tau=1}^{T-1}(1-\frac{\tau}{T-1})\beta(\tau)^{\frac{\eta}{4+\eta}}\{E[W_{t,(i,j)}(h)]^{2+\eta/2}\}^{\frac{1}{2+\eta/2}}\{E[W_{t+\tau,(i,j)}(h)]^{2+\eta/2}\}^{\frac{1}{2+\eta/2}}$$

$$= O(\frac{l^2}{T}).$$

Thus by Chebyshev's inequality, we have the result holds immediately. □

Lemma B.0.5. *Under Assumptions 3.2.3-3.3.4, $Q_T(h,\hat{W}) \rightarrow^p Q(h^o, W)$ uniformly for $h \in \Theta_l$.*

Proof of Lemma B.0.5

$Q_T(h,\hat{W}) - Q(h^o,W) = \hat{m}_T(h)'\hat{W}^{-1}\hat{m}_T(h) - Em^l(U_t,h^o)'W^{-1}Em^l(U_t,h^o)$

$= ([\hat{m}_T(h) - E\hat{m}_T(h) + E\hat{m}_T(h) - Em^l(U_t,h^o) + Em^l(U_t,h^o)]'\hat{W}^{-1}$

$[[\hat{m}_T(h) - E\hat{m}_T(h) + E\hat{m}_T(h) - Em^l(U_t,h^o) + Em^l(U_t,h^o)] - Em^l(U_t,h^o)'W^{-1}Em^l(U_t,h^o)$

$= [\hat{m}_T(h) - E\hat{m}_T(h)]'\hat{W}^{-1}[\hat{m}_T(h) - E\hat{m}_T(h)] + 2[\hat{m}_T(h) - E\hat{m}_T(h)]'\hat{W}^{-1}[E\hat{m}_T(h) - Em^l(U_t,h^o)]$

$+ 2[\hat{m}_T(h) - E\hat{m}_T(h)]'\hat{W}^{-1}Em^l(U_t,h^o) + [E\hat{m}_T(h) - Em^l(U_t,h^o)]'\hat{W}^{-1}[E\hat{m}_T(h) - Em^l(U_t,h^o)]$

$+ 2[E\hat{m}_T(h) - Em^l(U_t,h^o)]'\hat{W}^{-1}Em^l(U_t,h^o) + Em^l(U_t,h^o)'(\hat{W}^{-1} - W^{-1})Em^l(U_t,h^o)$

$= A_1 + A_2 + A_3 + A_4 + A_5 + A_6.$

In the following step, we are going to prove that A_i is tight and pointwise convergence for all $h \in \Theta_p$ for all $i = 1, \cdots, 6$. First, we start with A_1. By Assumption

$$\begin{aligned}
|A_1| &= |[\hat{m}_T(h) - E\hat{m}_T(h)]'\hat{W}^{-1}[\hat{m}_T(h) - E\hat{m}_T(h)]| \\
&\le \frac{1}{\lambda_{\min}(\hat{W})}[\hat{m}_T(h) - E\hat{m}_T(h)]'[\hat{m}_T(h) - E\hat{m}_T(h)] \\
&= \frac{1}{\lambda_{\min}(\hat{W})}\sum_{j=1}^{l}[\frac{1}{T}\sum_{t=1}^{T}m_j(U_t,h) - \frac{1}{T}E\sum_{t=1}^{T}m_j(U_t,h)]^2 \\
&= \frac{1}{\lambda_{\min}(\hat{W})}\frac{l}{\sqrt{T}}\frac{1}{l}\zeta_T'\zeta_T
\end{aligned}$$

By Lemma B.0.2, $\frac{1}{\lambda_{\min}(\hat{W})}\frac{l}{\sqrt{T}}$ is asymptotically bounded by $\frac{1}{\lambda_{\min}(W)}\frac{l}{\sqrt{T}}$ which goes to zero by assumption. In addition, we proved that $\frac{1}{l}\zeta_T'\zeta_T$ is tight. Thus A_1 is tight. What's more, by Lemma 1,

$$A_1 \le \frac{1}{\lambda_{\min}(\hat{W})}\frac{l}{\sqrt{T}}\frac{1}{l}(\zeta_T'\zeta_T - E\zeta_T'\zeta_T) + \frac{1}{\lambda_{\min}(\hat{W})}\frac{l}{\sqrt{T}}\frac{1}{l}E\zeta_T'\zeta_T \to^p 0$$

Thus $A_1 \to^p 0$ uniformly over $h \in \Theta$.

Next, we want to show that $A_2 \to^p 0$ uniformly. By Lemma B.0.1, for each $j = 1, \cdots, l$, $|E\hat{m}_{Tj}(h) - Em_j(h^o)| \leq |Ec_j(U_t)| ||h - h^o||$. Thus,

$$
\begin{aligned}
|A_2| &= 2|[\hat{m}_T(h) - E\hat{m}_T(h)]'\hat{W}^{-1}[E\hat{m}_T(h) - Em^l(U_t, h^o)]| \\
&\leq \frac{2}{\lambda_{\min}(\hat{W})}||h - h^o|||[\hat{m}_T(h) - E\hat{m}_T(h)]'Ec(U_t)| \\
&\leq \frac{O(p^{-s})}{\lambda_{\min}(\hat{W})}\frac{l}{\sqrt{T}}\frac{1}{l}\sum_{j=1}^{l}[Ec_j(U_t)]\zeta_{Tj}(h) \\
&\leq \frac{O(p^{-s})}{\lambda_{\min}(\hat{W})}\frac{l}{\sqrt{T}}[\frac{1}{l}\sum_{k=1}^{l}Ec_k^2(U_t)]^{\frac{1}{2}}[\frac{1}{l}\zeta_T(h)'\zeta_T(h)]^{\frac{1}{2}} \to 0.
\end{aligned}
$$

The convergence in probability follows immediately from $A_1 \to^p 0$ and assumptions $\frac{l}{\lambda_{\min}(W)\sqrt{T}} \to 0$.

Similarly, it is easy to prove that $A_3 \to^p 0$ and is tight.

$$
|A_3| = 2[\hat{m}_T(h) - E\hat{m}_T(h)]'\hat{W}^{-1}Em^l(U_t, h^o) \leq \frac{2}{\lambda_{\min}(\hat{W})}\frac{l}{\sqrt{T}}\frac{1}{l}\sum_{j=1}^{l}Em_j(U_t, h^o)\zeta_{Tj}(h)
$$

$$
\leq \frac{2}{\lambda_{\min}(\hat{W})}\frac{l}{\sqrt{T}}[\frac{1}{l}\sum_{j=1}^{l}Em_j^2(U_t, h^o)]^{\frac{1}{2}}[\frac{1}{l}\zeta_T(h)'\zeta_T(h)]^{\frac{1}{2}} \to 0.
$$

Under Lemma B.0.1, for each $j = 1, \cdots, l$, $|E\hat{m}_{Tj}(h) - Em_j(U_t, h^o) \leq |[Ec_j(U_t)]| \cdot ||h - h^o||$, then

$$
\begin{aligned}
|A_4| &= |E\hat{m}_T(h) - Em^l(U_t, h^o)]'\hat{W}^{-1}[E\hat{m}_T(h) - Em^l(U_t, h^o)]| \\
&\leq \frac{1}{\lambda_{\min}(\hat{W})}\sum_{j=1}^{l}[Ec_j(U_t)]^2||h - h^o||^2 = \mathbb{O}_p(\frac{lp^{-2s}}{\lambda_{\min}(W)}) = \mathbb{o}_p(1).
\end{aligned}
$$

Therefore, $|A_4|$ is uniformly convergent to 0 for $h \in \Theta$. Analogously, we can prove

$A_5 \to^p 0$ uniformly. By the Cauchy-Schwarz inequality, we have

$$|A_5| = |2[E\hat{m}_T(h) - Em^l(U_t, h^o)]'\hat{W}^{-1}Em^l(U_t, h^o)| \leq \frac{2}{\lambda_{\min}(\hat{W})}||h - h^o|| \sum_{j=1}^{l} |Ec_j(U_t)Em_j(U_t, h^o)|$$

$$\leq \frac{\mathbb{O}(lp^{-s})}{\lambda_{\min}(\hat{W})}\{\frac{1}{l}\sum_{j=1}^{l}[Ec_j(U_t)]^2\}^{\frac{1}{2}}\{\frac{1}{l}\sum_{k=1}^{l}[Em_j(h^o)]^2\}^{\frac{1}{2}} = \mathbb{O}_p\big(\frac{lp^{-s}}{\lambda_{\min}(\hat{W})}\big) = \mathbb{o}_p(1).$$

Lastly, we show that $A_6 \to^p 0$ uniformly for $h \in \Theta$. First, we prove that A_6 is tight and pointwise convergence. We have,

$$\begin{aligned}
|A_6| &= |Em^l(U_t, h^o)'[\hat{W}^{-1}(h) - W^{-1}(h^o)]Em^l(U_t, h^o)| \\
&= |Em^l(U_t, h^o)'\hat{W}^{-1}(h)[\hat{W}(h) - W(h^o)]W^{-1}(h^o)]Em^l(U_t, h^o)| \\
&\leq |Em^l(U_t, h^o)'\hat{W}^{-1}(h)[\hat{W}(h) - \hat{W}(h^o)]W^{-1}(h^o)Em^l(U_t, h^o)| \\
&+ |Em^l(U_t, h^o)'\hat{W}^{-1}(h)[\hat{W}(h^o) - W(h^o)]W^{-1}(h^o)Em^l(U_t, h^o)| \\
&= A_6^{(1)} + A_6^{(2)}.
\end{aligned}$$

Considering the definition of the optimal weighting matrix $W(h^o) = E[m^l(U_t, h^o)m^l(U_t, h^o)']$, we have:

$$\begin{aligned}
A_6^{(1)} &= |Em^l(U_t, h^o)'\hat{W}^{-1}(h)[\hat{W}(h) - \hat{W}(h^o)]W^{-1}(h^o)Em^l(U_t, h^o)| \\
&= |tr\{\hat{W}^{-1}(h)[\hat{W}(h) - \hat{W}(h^o)]W^{-1}(h^o)Em^l(U_t, h^o)Em^l(U_t, h^o)'\}| \\
&= O_p(\frac{lp^{-s}}{\lambda_{\min}W}) = o_p(1).
\end{aligned}$$

Because $W - Em^l(U_t, h^o)m^l(U_t, h^o)'$ is positive definite, it is useful to note that for any real valued vector $c \in R^l$, such that $c'c = 1$, we have

$$1 - c'[W^{-1}Em^l(U_t, h^o)m^l(U_t, h^o)']c = c'[I - W^{-1}Em^l(U_t, h^o)m^l(U_t, h^o)']c \geq 0.$$

It implies that

$$\lambda_{\max}[W^{-1}Em^l(U_t,h^o)m^l(U_t,h^o)'] \leq 1.$$

Using Lemma B.0.4 , we have

$$\begin{aligned}
A_6^{(2)} &= |Em^l(U_t,h^o)'\hat{W}^{-1}(h^o)[\hat{W}(h^o)-W(h^o)]W^{-1}(h^o)Em^l(U_t,h^o)| \\
&\leq \frac{1}{\lambda_{\min}\hat{W}}|tr\{[\hat{W}(h^o)-W(h^o)]W^{-1}(h^o)Em^l(U_t,h^o)Em^l(U_t,h^o)'\}| \\
&\leq \frac{1}{\lambda_{\min}W}\sqrt{tr[\hat{W}(h^o)-W(h^o)]^2} \\
&= O_p(\frac{l}{\lambda_{\min}(W)\sqrt{T}}) = o_p(1).
\end{aligned}$$

Therefore, this establish the uniform convergence of $Q_T(h,\hat{W}) \to Q(h^o,W)$ for all $h \in \Theta$. □

Lemma B.0.6. *Under Assumptions 3.2.3-3.3.3,* $||h^l(X_t) - h^o(X_t)||_{d,w} = o_p(1)$ *for* $h_l \in \Theta$.

Proof of Lemma B.0.6 From Lemma B.0.5, the result follows immediately from the Theorem 2.7 in Newey (1994) . □

Proposition 2. *Suppose that* $\bar{Q}(h_l,W(h_l))$ *is twice differentiable with respect to* γ^l. *Assume For fixed constants* C, *for every* T, *and for every sufficiently small* $\delta > 0$, *and* $h_l \in \Theta_l$,

$$\sup_{d(h_l,h_{lo})<\delta} \bar{Q}(h_l,\hat{W}(h_l)) - \bar{Q}(h_{lo},W(h_{lo})) \geq C\delta^2.$$

Proof of Proposition 2 It is easy to see that $m_k(U_t,h_i)$ is differentiable at γ_i, where $h_i = \phi(X_t)\gamma_i^l$. Thus $m_j(U_t,h_l) - m_j(U_t,h_{lo}) \approx (\gamma^l - \gamma^{lo})'\dot{m}_j(U_t,h_{lo})$ for $j = 1,\cdots,l$, where $\dot{m}_j(U_t, h_{lo}) = \frac{\partial m_j(U_t,h)}{\partial \gamma^l}|_{\gamma^l=\gamma^{lo}}$. Therefore, applying a quadratic approximation, we have:

$$\bar{Q}_T\big(h_1, \hat{W}(h_1)\big) - \bar{Q}_T\big(h_2, \hat{W}(h_2)\big) = \frac{1}{2}(\gamma^l - \gamma^{lo})'\phi(X_t)V\phi(X_t)'(\gamma^l - \gamma^{lo}) + R(h_l)$$

$$\geq C(h_l - h_{lo})^2 + \epsilon(h_l - h_{lo})^2 \geq C\delta^2.$$

The inequality holds because V is positive given h_{lo} uniquely minimizes $\bar{Q}(h_l, W(h_l))$ which is twice differentiable. □

Proposition 3. *Let $r_T = \tilde{r}_T\sqrt{T}$, where $\tilde{\tau} = \frac{\lambda_{\min}(W)}{l}$. Suppose h_{lo} uniquely minimizes $\hat{Q}_T(h_l)$ for $h_l \in \Theta_l$. For all $h \in \Theta_l$ and $\delta > 0$,*

$$E \sup_{||h-h_{lo}||\leq\delta} |r_T M_T(h, \hat{W}(h)) - r_T M_T(h_{lo}, W(h_{lo}))| \lesssim \delta.$$

Proof of Proposition 3 It is helpful to prove this lemma by considering the the generalized likelihood estimator (GEL) $\hat{h}_{GEL} = \arg\min_{h\in\Theta_l} \sup \sum_{t=1}^{T} \rho(\lambda' m_i(h))$. Using Theorem 2.1 by Newey and Smith (2004), GEL coincides with CUE when $\rho(v)$ is quadratic, that $\hat{h}_{GMM} = \hat{h}_{GEL}$. Let

$$l(h, \hat{W}(h)) = [\hat{m}_T(h)'\hat{W}^{-1}(h)m(U_t, h) - \frac{1}{2}\hat{m}_T(h)'\hat{W}^{-1}(h)m(U_t, h)m(U_t, h)'\hat{W}^{-1}(h)\hat{m}_T(h)].$$

Thus we can equivalently express the objective function as a special case of GEL as follows:

$$\hat{P}_T(h) = \frac{1}{T}\sum_{t=1}^{T} l(h, \hat{W}(h)).$$

And let $\tilde{M}_T(h, \hat{W}(h)) = \hat{P}_T(h) - E\hat{P}_T(h)$. It suffices to prove

$$E \sup_{||h-h_{lo}||\leq\delta} |r_T \tilde{M}_T(h, \hat{W}(h)) - r_T \tilde{M}_T(h_{lo}, W(h_{lo}))| \lesssim \delta.$$

Specifically, we have

$$l(h_1, \hat{W}(h_1)) - l(h_2, \hat{W}(h_2)) = \hat{m}_T(h_1)'\hat{W}^{-1}(h_1)m^l(U_t, h_1) - \hat{m}_T(h_2)'\hat{W}^{-1}(h_2)m^l(U_t, h_2)$$
$$+ \frac{1}{2}[\hat{m}_T(h_2)'\hat{W}^{-1}(h_2)m^l(U_t, h_2)m^l(U_t, h_2)'\hat{W}^{-1}(h_2)\hat{m}_T(h_2)$$
$$- \hat{m}_T(h_1)'\hat{W}^{-1}(h_1)m^l(U_t, h_1)m^l(U_t, h_1)'\hat{W}^{-1}(h_1)\hat{m}_T(h_1)] = B_1 + B_2$$

where

$$B_1 = [\hat{m}_T(h_1) - \hat{m}_T(h_2)]'\hat{W}^{-1}(h_1)m^l(U_t, h_1) + \hat{m}_T(h_2)\hat{W}^{-1}(h_2)[m^l(U_t, h_1) - m^l(U_t, h_2)]$$
$$+ \hat{m}_T(h_2)'[\hat{W}^{-1}(h_1) - \hat{W}^{-1}(h_2)]m^l(U_t, h_1) = B_1^1 + B_1^2 + B_1^3.$$

We prove that $|B_1^i| \lesssim \frac{l}{\lambda_{\min}(W)}||h_1 - h_2||$ for all $i = 1, 2, 3$. Applying the Cauchy-Schwarz inequality and Lemma B.0.1, we have

$$\begin{aligned}
&E|B_1^1| \equiv |[\hat{m}_T(h_1) - \hat{m}_T(h_2)]'\hat{W}^{-1}(h_1)m^l(U_t, h_1)| \\
&\leq \frac{l}{\lambda_{\min}(W)}\sqrt{\frac{1}{l}\sum_{j=1}^{l} E[\hat{m}_{Tj}(h_1) - \hat{m}_{Tj}(h_2)]^2} \cdot \sqrt{\frac{1}{l}\sum_{j=1}^{l} Em_j^2(U_t, h_1)} \\
&\leq O_p(1)\frac{l}{\lambda_{\min}(W)}\sup_{1\leq j\leq l}[E|\frac{1}{T}\sum_{t=1}^{T} c_j(U_t)|^2]^{\frac{1}{2}}||h_1 - h_2|| \\
&\lesssim \frac{l}{\lambda_{\min}(W)}||h_1 - h_2||.
\end{aligned}$$

Then, similarly using the Cauchy-Schwarz inequality and Lemma B.0.1, we have

$$\begin{aligned}
&E|B_1^2| = E|\hat{m}_T(h_2)'\hat{W}^{-1}(h_2)[m(U_t, h_1) - m(U_t, h_2)]| \\
&\leq \frac{l}{\lambda_{\min}(W)}\sqrt{\frac{1}{l}\sum_{j=1}^{l} E[\frac{1}{T}\sum_{t=1}^{T} m_j(U_t, h_2)]^2}\sqrt{\frac{1}{l}\sum_{j=1}^{l} Ec_j^2(U_t)}||h_1 - h_2|| \\
&\lesssim \frac{l}{\lambda_{\min}(W)}||h_1 - h_2||.
\end{aligned}$$

Finally, using Lemma B.0.1 and the property that $tr(AB) \leq \sqrt{tr(A^2)tr(B^2)}$, we have

$$
\begin{aligned}
E|B_1^3| &= E|\hat{m}_T(h_2)'[\hat{W}^{-1}(h_1) - \hat{W}^{-1}(h_2)]m(U_t, h_1)| \\
&= E|\hat{m}_T(h_2)'\hat{W}^{-1}(h_1)[\hat{W}(h_1) - \hat{W}(h_2)]\hat{W}^{-1}(h_2)m(U_t, h_1)| \\
&\leq \frac{1}{\lambda_{\min}(W)} Etr|[\hat{W}(h_1) - \hat{W}(h_2)]\hat{W}^{-\frac{1}{2}}(h_1)m(U_t, h_1)\hat{m}_T(h_2)'\hat{W}^{-\frac{1}{2}}(h_2)| \\
&\lesssim \frac{1}{\lambda_{\min}(W)} E||\hat{W}(h_1) - \hat{W}(h_2)||E||m(U_t, h_2)W^{-\frac{1}{2}}(h_2)|| + o_p(1) \\
&\lesssim \frac{l}{\lambda_{\min}(W)}||h_1 - h_2||.
\end{aligned}
$$

Next, we prove that $|B_2| = O_p(\frac{l}{\lambda_{\min}(W)}||h_1 - h_2||)$. Let $\hat{R}(U_t, h) = \hat{W}^{-1}(h)m(U_t, h)m(U_t, h)'$. It is helpful to show that $\lambda_{\max}(\hat{R}(U_t, h)) \leq 1$, as $T \to \infty$. Using this fact, we have

$$
\begin{aligned}
2B_2 &= [\hat{m}_T(h_2)'\hat{W}^{-1}(h_2)m(U_t, h_2)m(U_t, h_2)'\hat{W}^{-1}(h_2)\hat{m}_T(h_2) \\
&\quad - \hat{m}_T(h_1)'\hat{W}^{-1}(h_1)m(U_t, h_1)m(U_t, h_1)'\hat{W}^{-1}(h_1)\hat{m}_T(h_1)] \\
&= [\hat{m}_T(h_2) - \hat{m}_T(h_1)]'\hat{R}(h_2)\hat{W}^{-1}(h_2)\hat{m}_T(h_2) + \hat{m}_T'(h_1)[\hat{R}(h_2) - \hat{R}(h_1)]\hat{W}^{-1}(h_2)\hat{m}_T(h_2) \\
&\quad + \hat{m}_T(h_1)\hat{R}(h_1)[\hat{W}^{-1}(h_2) - \hat{W}^{-1}(h_1)]\hat{m}_T(h_2) + \hat{m}_T(h_1)\hat{R}(h_1)\hat{W}^{-1}(h_1)[\hat{m}_T(h_2) - \hat{m}_T(h_1)] \\
&= B_2^1 + B_2^2 + B_2^3 + B_2^4.
\end{aligned}
$$

Using the triangular inequality, we have $2E|B_2| \leq E|B_2^1| + E|B_2^2| + E|B_2^3| + E|B_2^4|$, and we show that $E|B_2^i| = O_p(\frac{l}{\lambda_{\min}(W)})||h_1 - h_2||$. First,

$$E|B_2^1| = E|[\hat{m}_T(h_2) - \hat{m}_T(h_1)]'\hat{W}^{-1}(h)m(U_t,h)m(U_t,h)'\hat{W}^{-1}(h_2)\hat{m}_T(h_2)|$$
$$\leq \frac{1}{\lambda_{\min}(W)}E|tr\big[m(U_t,h)m(U_t,h)'\hat{W}^{-1}(h)\hat{m}_T(h_2)[\hat{m}_T(h_2) - \hat{m}_T(h_1)]'\big]|$$
$$\leq \frac{l}{\lambda_{\min}(W)}E|\lambda_{\max}\big(m(U_t,h)m(U_t,h)'\hat{W}^{-1}(h)\big)\big(\sup_{1\leq j\leq l}\frac{1}{T}\sum_{t=1}^{T}m_{Tj}(U_t,h_2)\frac{1}{T}\sum_{t=1}^{T}c_j(U_t)||h_2 - h_1||\big)|$$
$$= O_p(1)\frac{l}{\lambda_{\min}(W)}||h_2 - h_1||.$$

By the Markov's inequality, we have $|B_2^1| = O_p(\frac{l}{\lambda_{\min}(W)}||h_1 - h_2||)$. Analogously, we are able to show that $|B_2^4| = o_p(\frac{1}{\lambda_{\min}(W)}||h_2 - h_1||)$. Applying Lemma B.0.1, we have

$$B_2^2 = \hat{m}_T(h_1)'\hat{W}^{-1}(h_2)[m(U_t,h_2)m(U_t,h_2)' - m(U_t,h_1)m(U_t,h_1)']\hat{W}^{-1}(h_2)\hat{m}_T(h_2)$$
$$+ \hat{m}_T(h_1)'\hat{W}^{-1}(h_2)[\hat{W}(h_1) - \hat{W}(h_2)]\hat{W}^{-1}(h_1)m(U_t,h_1)m(U_t,h_1)'\hat{W}^{-1}(h_2)\hat{m}_T(h_2)$$
$$= B_2^{2,1} + B_2^{2,2}.$$

Specifically, using the triangular inequality, we have

$$E|B_2^{2,1}| = Etr|\hat{m}_T(h_1)'\hat{W}^{-1}(h_1)[m(U_t,h_2)m(U_t,h_2)' - m(U_t,h_1)m(U_t,h_1)']\hat{W}^{-1}(h_2)\hat{m}_T(h_2)|$$
$$+ Etr|\hat{m}_T(h_1)'[\hat{W}^{-1}(h_2) - \hat{W}^{-1}(h_1)][m(U_t,h_2)m(U_t,h_2)'$$
$$- m(U_t,h_1)m(U_t,h_1)']\hat{W}^{-1}(h_2)\hat{m}_T(h_2)|$$
$$= EB_2^{2,1,1} + EB_2^{2,1,2}.$$

Recall that $\lambda_{\max}(A) = a'Aa$, from a vector $||a|| = 1$. Applying the property of trace and Lemma B.0.3, we have

$$
\begin{aligned}
&EB_2^{2,1,1} = Etr|\hat{m}_T(h_1)'\hat{W}^{-1}(h_1)[m(U_t,h_2)m(U_t,h_2)' - m(U_t,h_1)m(U_t,h_1)']\hat{W}^{-1}(h_2)\hat{m}_T(h_2)| \\
&\le \frac{1}{\lambda_{\min}(W)}E|tr\{[m(U_t,h_2)m(U_t,h_2)' - m(U_t,h_1)m(U_t,h_1)']\hat{W}^{-\frac{1}{2}}(h_2)\hat{m}_T(h_2)\hat{m}_T(h_1)'\hat{W}^{-\frac{1}{2}}(h_1)\}| \\
&\le \frac{1}{\lambda_{\min}(W)}E|\lambda_{\max}[\hat{W}^{-\frac{1}{2}}(h_2)\hat{m}_T(h_2)\hat{m}_T(h_1)'\hat{W}^{-\frac{1}{2}}(h_1)]|tr[m(U_t,h_2)m(U_t,h_2)' - m(U_t,h_1)m(U_t,h_1)']| \\
&\le \frac{l}{\lambda_{\min}(W)}E|\lambda_{\max}[m(U_t,h_2)m(U_t,h_2)' - m(U_t,h_1)m(U_t,h_1)']| \\
&= O_p(\frac{l}{\lambda_{\min}(W)})||h_2 - h_1||.
\end{aligned}
$$

In addition, it is easy to show that $m(U_t,h)m(U_t,h)'$ is a positive definite matrix, whose eigenvalues are all positive.

$$
\begin{aligned}
&EB_2^{2,1,2} = Etr|\hat{m}_T(h_1)'[\hat{W}^{-1}(h_2) - \hat{W}^{-1}(h_1)][m(U_t,h_2)m(U_t,h_2)' - m(U_t,h_1)m(U_t,h_1)']\hat{W}^{-1}(h_2)\hat{m}_T(h_2 \\
&\le \frac{1}{\lambda_{\min}(W)}Etr|\hat{m}_T(h_1)'\hat{W}^{-\frac{1}{2}}(h_1)[\hat{W}(h_1) - \hat{W}(h_2)]\hat{W}^{-1}(h_2)[m(U_t,h_2)m(U_t,h_2)']\hat{W}^{-\frac{1}{2}}(h_2)\hat{m}_T(h_2)| \\
&+ \frac{1}{\lambda_{\min}(W)}Etr|\hat{m}_T(h_1)'\hat{W}^{-\frac{1}{2}}(h_1)[\hat{W}(h_1) - \hat{W}(h_2)]\hat{W}^{-1}(h_2)[m(U_t,h_1)m(U_t,h_1)']\hat{W}^{-\frac{1}{2}}(h_2)\hat{m}_T(h_2)| \\
&\le \frac{1}{\lambda_{\min}(W)}E\{\lambda_{\max}[\hat{W}^{-\frac{1}{2}}(h_2)\hat{m}_T(h_2)\hat{m}_T(h_1)'\hat{W}^{-\frac{1}{2}}(h_1)]tr|\hat{W}^{-1}(h_2)m(U_t,h_2)m(U_t,h_2)'|\}||h_2 - h_1|| \\
&+ \frac{1}{\lambda_{\min}(W)}E\{tr[\hat{W}^{-\frac{1}{2}}(h_2)\hat{m}_T(h_2)\hat{m}_T(h_1)'\hat{W}^{-\frac{1}{2}}(h_1)]\lambda_{\max}[\frac{\hat{W}^{-1}(h_2)}{\lambda_{\max}[W^{-1}(h_1)]}\hat{W}^{-1}(h_1)m(U_t,h_1)m(U_t,h_1)']\}| \\
&\le \frac{1}{\lambda_{\min}(W)}||h_2 - h_1||\{Etr[\hat{W}^{-1}(h_2)m(U_t,h_2)m(U_t,h_2)'] + lE\lambda_{\max}[\hat{W}^{-1}(h_1)m(U_t,h_1)m(U_t,h_1)']\} \\
&= \frac{l}{\lambda_{\min}(W)}||h_2 - h_1||.
\end{aligned}
$$

Also, applying the property that $tr(AB) \leq tr(A)tr(B)$, we have

$$E|B_2^3| = E|\hat{m}_T(h_1)'\hat{W}^{-1}(h_1)[m(U_t,h_1)m(U_t,h_1)'][\hat{W}^{-1}(h_2) - \hat{W}^{-1}(h_1)]\hat{m}_T(h_2)|$$
$$= O_p(1)\frac{1}{\lambda_{\min}(W)}||h_1 - h_2||E\{tr[\hat{W}^{-\frac{1}{2}}(h_1)\hat{W}^{-\frac{1}{2}}(h_2)\hat{m}_T(h_1)'\hat{m}_T(h_2)\hat{W}^{-1}(h_1)m(U_t,h_1)m(U_t,h_1)'\}$$
$$= O_p(1)\frac{1}{\lambda_{\min}(W)}||h_1 - h_2||trE[\hat{W}^{-1}(h_1)m(U_t,h_1)m(U_t,h_1)']$$
$$= O_p(1)\frac{l}{\lambda_{\min}(W)}||h_2 - h_1||.$$

We prove rest part of this lemma with some modifications of the proof used in Collollary 5.53 by Sara Van de Geer (1998). For all $h_1, h_2 \in \Theta_l$, First, by Assumption 3.3.3,

$$\frac{\lambda_{\min}(W)}{l}|l(h_1,\hat{W}(h_1)) - l(h_2,\hat{W}(h_2))| \lesssim ||h_2 - h_1||.$$

Define the class of functions $\mathbb{F} = \{\tilde{r}_T l(h_1,\hat{W}(h_1)) - \tilde{r}_T l(h_2,\hat{W}(h_2)) : ||h_1 - h_2|| \leq \delta\}$. This class of functions have an envelope $F = \delta$. We apply Corollary 19.35 by Sara Van de Geer (1998), whence

$$E^* \sup_{||h-h_{lo}||\leq\delta} |r_T\tilde{M}_T(h,\hat{W}(h)) - r_T\tilde{M}_T(h^*)| \lesssim \int_0^\delta \sqrt{logN_{[]}(\epsilon,\mathbb{F},L_2(P))}d\epsilon.$$

Using $h \in L_w^2$ and Corollary 2.6 in Van de Geer (2000), we have $logN_{[]}(\epsilon,\mathbb{F},L_2(P)) \leq llog(\frac{4R+\epsilon}{\epsilon})$, where the integral is bounded by a multiple of δ. □

Proof of Theorem 3.3.2 We prove this theorem by using the Proposition 10 in Han and Phillips (2006), Lemma (B.0.3) and Lemma (B.0.5). We first establish $||h_l - h_{lo}|| = O_p(r_T^{-1}) = O_p(\frac{l}{\lambda_{\min}(W)\sqrt{T}})$.

For each T, the parameter space h minus the point h^o can be partitioned into shells, $s_{j,T} = \{h \in \Theta_l : 2^{j-1} < r_T d(h - h_{lo}) \leq 2^j\}$, with j ranging over the integers.

If $r_T d(h, h_{lo}) > 2^M$ for a given integer M, then $\hat{h}_l$ will fall in one of the shells $s_{j,T}$ with $j \geq M$.

Suppose $\hat{h}_l$ maximizes the map $h \to -\hat{Q}_T(h, \hat{W}(h))$ up to a variable $R_T = O_p(r_T^{-2})$. By the property of $\hat{h}_{lT}$, we conclude that for every $\epsilon > 0$,

$$P^\star(r_T d(h - h_{lo}) > 2^M) \leq \sum_{j \geq M, 2^j \leq \epsilon r_T} P^\star\{- \sup_{h \in s_{j,T}} [\hat{Q}_T(h, \hat{W}(h)) - \hat{Q}_T(h_{lo}, \hat{W}(h_{lo})] \geq -\frac{K}{r_T^2}\}$$
$$+ P[2d(\hat{h}_l - h_{lo}) \geq \epsilon] + P(r_T^2 R_T \geq K).$$

Because we proved $\hat{h}_l$ is consistent for h_{lo}, the second probability on the right is equal to 0 as $T \to \infty$ for every fixed $\epsilon > 0$. The third probability can be made arbitrarily small uniformly in T by choice of K. We choose ϵ small enough and $\delta < \epsilon$. By Proposition 2, for every j involved in the sum, we have

$$-[\sup_{h \in s_{j,T}} \bar{Q}_T\big(h, \hat{W}(h)\big) - \bar{Q}_T\big(h_{lo}, \hat{W}(h_{lo})\big)] \leq -C \frac{2^{2j-2}}{r_T^2}.$$

Now, for $\frac{1}{2} C r_T 2^{2(M-1)} \geq K$, by the Chebyshev's inequality and Proposition 3, we can bound the series as follows:

$$\sum_{j\geq M, 2^j\leq \epsilon r_T} P^\star\{-\sup_{h\in s_{j,T}}[\hat{Q}_T(h,\hat{W}(h)) - \hat{Q}_T(h_{lo},\hat{W}(h_{lo})] \geq -\frac{K}{r_T^2}\}$$

$$\leq \sum_{j\geq M, 2^j\leq \epsilon r_T} P^\star\{-\sup_{h\in s_{j,T}} r_T[\hat{Q}_T(h,\hat{W}(h)) - \hat{Q}_T(h_{lo},\hat{W}(h_{lo})]$$

$$+ r_T[\bar{Q}_T(h,\hat{W}(h)) - \bar{Q}_T(h_{lo},\hat{W}(h_{lo})] \geq C\frac{2^{2(j-1)}}{2r_T}\}$$

$$= \sum_{j\geq M, 2^j\leq \epsilon r_T} P^\star\{\sup_{j\in s_{i,T}} |r_T M_T(h,\hat{W}(h)) - r_T M_T(h_{lo},\hat{W}(h_{lo}))| \geq C\frac{2^{2(j-1)}}{2r_T}\}$$

$$\leq \sum_{j\geq M, 2^j\leq \epsilon r_T} \frac{E\sup_{j\in s_{i,T}} |r_T M_T(h,\hat{W}(h)) - r_T M_T(h_{lo},\hat{W}(h_{lo}))|}{C\frac{2^{2(j-1)}}{2r_T}}$$

$$\leq \sum_{j\geq M, 2^j\leq \epsilon r_T} \frac{\frac{2^j}{r_T}}{C\frac{2^{2(j-1)}}{2r_T}} = \sum_{j\geq M, 2^j\leq \epsilon r_T} 4\cdot 2^{-j} \to 0 \text{ as } M\to\infty.$$

Finally, we establish the result by applying the triangular inequality that for any given $x \in \mathbf{X}$,

$$||\hat{h}_l - h^o||_{d,w} \leq ||h_l - h_{lo}|| + ||h_{lo} - h^o|| = O_p(\frac{l}{\lambda_{\min}(W)\sqrt{T}} + p^{-s}).$$

□

Lemma B.0.7. *Suppose Assumption 3.3.4 hold. Then for all $h_l \in \Theta_l$ in a neighborhood of γ_{lo}, we have 1) $\sqrt{\frac{T}{l}}\sup_{\gamma_l\in\Theta_l}||\hat{m}_T(h_l)|| = O_p(1)$; 2) $\sqrt{\frac{T}{l}}\sup_{\gamma_l\in\Theta_l}||\partial\hat{m}_T(h_l)/\partial\gamma_j|| = O_p(1)$; 3) $\frac{\sqrt{T}}{l}\sup_{\gamma_l\in\Theta_l}||\partial\hat{m}_T(h_l)/\partial\gamma|| = O_p(1)$; 4) $\frac{\sqrt{T}}{l}\sup_{\gamma_l\in\Theta_l}||\partial^2\hat{m}_T(h_l)/\partial\gamma\partial\gamma'|| = O_p(1)$. In order words, all these terms are uniformly bounded by a constant.*

Proof of Lemma B.0.7 For part 1), using Theorem 5.5 in Roussas and Ioannides inequality, we have

$$E \sup_{h_l \in \Theta_l} ||\hat{m}_T(h_l)||^2 \leq \frac{1}{T-1} E|m_t(h_l)' m_t(h_l)| + \frac{1}{T}\sum_{\tau=2}^{T}(1-\frac{\tau}{T})\beta^{\frac{\delta}{1+\delta}}[E|m_1(h_l)'m_\tau(h_l)|^{1+\delta}]^{\frac{1}{1+\delta}}$$
$$\leq \frac{l}{T}\sup_{1\leq j\leq l} Em_j^2(U_t,h_l) + \frac{l}{T}[Em_j^{2+2\delta}(U_t,h_l)]^{\frac{1}{1+\delta}} = O_p(\frac{l}{T}).$$

Thus, by the Markov's inequality, we have $\sqrt{\frac{T}{l}}\sup_{\gamma_l \in \Theta_l} ||\hat{m}_T(h_l)|| = O_p(1)$.

By Assumption 3.3.4, Part 2) follows immediately given $E(\frac{\partial m_j(U_t,h_l)}{\partial \gamma_i})^{2+2\delta}$ is uniformly bounded for all $i, j = 1, \cdots, l$. Also, noting that $\sup_{h_l\in\Theta_l} ||\partial \hat{m}_T(h_l)/\partial \gamma^l||$ is an $l \times l$ matrix with $l \to \infty$ as $T \to \infty$, we have

$$E \sup_{h_l \in \Theta_l} ||\partial \hat{m}_T(h_l)/\partial\gamma^l||^2 \leq \sum_{i=1}^{l}\sum_{j=1}^{l} E[\frac{1}{T}\sum_{t=1}^{T}\frac{\partial m_j(U_t,h_l)}{\partial \gamma_i}]^2$$
$$\leq \frac{l^2}{T}\sup_{1\leq i,j\leq l} E[\frac{\partial m_j(U_t,h_l)}{\partial \gamma_i}]^2 + \frac{l^2}{T}[E|\frac{\partial m_j(U_t,h_l)}{\partial \gamma_i}|^{2+2\delta}]^{\frac{1}{1+\delta}} = O_p(\frac{l^2}{T}).$$

Therefore, by the Markov's inequality, we have $\frac{\sqrt{T}}{l}\sup_{\gamma_l\in\Theta_l} ||\partial \hat{m}_T(h_l)/\partial\gamma|| = O_p(1)$.

In part 3), we have for all $h_l \in \Theta_l$, $\partial^2 \hat{m}_T(h_l)/\partial\gamma^l\partial\gamma^{l'}$ is an $l\times l$ matrix. Under Assumption 3.3.4, we have $E|\frac{\partial^2 m_j(U_t,h_l)}{\partial\gamma_i\partial\gamma_j}|^{2+2\delta} \leq \Delta < \infty$ for all $i, j = 1, \cdots, l$. Therefore, it follows immediately that

$$E \sup_{\gamma_l \in \Theta_l} ||\partial \hat{m}_T^2(h_l)/\partial\gamma^l\partial\gamma^{l'}||^2 \leq \sum_{i=1}^{l}\sum_{j=1}^{l} E||\frac{1}{T}\sum_{t=1}^{T}\frac{\partial m_j(U_t,h_l)}{\partial \gamma_i\partial\gamma_j}||^2$$
$$\leq \frac{l^2}{T}\sup_{1\leq i,j\leq l} E[\frac{\partial m_j(U_t,h_l)}{\partial \gamma_i\partial\gamma_j}]^2 + \frac{l^2}{T}[E|\frac{\partial m_j(U_t,h_l)}{\partial \gamma_i\partial\gamma_j}|^{2+2\delta}]^{\frac{1}{1+\delta}} = O_p(\frac{l^2}{T}).$$

Thus, it is immediate that $\frac{\sqrt{T}}{l}\sup_{\gamma_l\in\Theta_l} ||\partial \hat{m}_T^2(h_l)/\partial\gamma^l\partial\gamma^{l'}|| = O_p(1)$. □

Lemma B.0.8. *Suppose Assumptions 3.2.3-3.3.4 hold. Suppose S is some non-stochastic vector in R^l and $T^{-\frac{\delta}{2}}[\frac{1}{\lambda_{\min} W}]^{1+\delta/2} \to 0$, then for $l \to \infty$ as $T \to \infty$, we* *have*

$$\sqrt{T} F S' G_l^{-1} D_0 W^{-1} \hat{m}_T(h) \xrightarrow{d} N(0,1), \tag{B.1}$$

where $D_0 = E[\frac{\partial m^l(U_t,h_l)}{\partial \gamma^l}]$, $G_l = D_0' W^{-1} D_0$, $\Sigma_l = E[m^l(U_t,h_l) m^l(U_t,h_l)']$, $V_l = S' G_l^{-1} D_0 \Sigma_l D_0' G_l^{-1} S$, *and* $F = V_l^{-\frac{1}{2}}$.

Proof of Lemma B.0.8 We first show that $V_l > 0$ is well-defined for sufficiently large T. Define $y_i = F S' G_l^{-1} D_0 W^{-1} e_i$, where e_i is a vector with the ith element equal to 1 and else 0. Let $Z_t = y m^l(U_t,h_l)$, therefore $\bar{Z}_t \equiv \frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_t = \sqrt{T} F S' G_l^{-1} D_0 W^{-1} \hat{m}_T(h)$. First, it is easy to show that Z_t is a MDS given $E_t(Z_t) = 0$. Second, recall Assumption that $m^l(U_t,h_l)$ is a measurable function of U_t and U_t is a mixing sequence with size $\alpha(T) = -r/(r-2)$, for some $r > 2$. We have Z_t is a mixing sequence of size $\alpha(T) = -r/(r-2)$ for some $r > 2$. Hence, we apply Brown's (1971) CLT theorem for martingale difference sequences. By the Minkowski and Markov inequalities, the properties of trace, and the condition that $\frac{\lambda^{2+\delta}(p)}{T^\delta} \to 0$, we have the following Lindeberg condition:

$$\begin{aligned}
& V_l^{-1} T^{-1} \sum_{t=1}^{T} E\big\{[\sum_{j=1}^{l} y_j m_j(U_t,h_l)]^2 I\{[\sum_{j=1}^{l} y_j m_j(U_t,h_l)]^2 \geq \epsilon T V_l\}\big\} \\
& \leq V_l^{-1} T^{-1} \sum_{t=1}^{T} (\epsilon T V_l)^{-\frac{\delta}{2}} E|\sum_{j=1}^{l} y_j m_j(U_t,h_l)|^{2+\delta} \leq F^{2+\delta} T^{-\frac{\delta}{2}} \{\sum_{j=1}^{l} y_j [E|m_j(U_t,h_l)|^{2+\delta}]^{\frac{1}{2+\delta}}\}^{2+\delta} \\
& \lesssim F^{2+\delta} T^{-\frac{\delta}{2}} |tr[S' G_l^{-1} D_0 D_0' G_l^{-1} S]|^{1+\frac{\delta}{2}} \\
& \lesssim T^{-\frac{\delta}{2}} [\frac{1}{\lambda_{\min} W}]^{1+\delta/2} F^{2+\delta} V_l^{1+\frac{\delta}{2}} \\
& \lesssim T^{-\frac{\delta}{2}} [\frac{1}{\lambda_{\min} W}]^{1+\delta/2} = o_p(1).
\end{aligned}$$

Next, we have

$$var(\bar{Z}_t) = F^2 S' G_l^{-1} D_0 W^{-1} E \frac{1}{T} \sum_{t=1}^{T} m(U_t, h) m(U_t, h)' W^{-1} D_0 G_l^{-1} AS = 1 > 0. \tag{B.2}$$

Therefore, we complete the CLT by showing that $\frac{1}{T}\sum_{t=1}^{T} Z_t^2 \to 1$. Using Lemma B.0.4, we have

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T} Z_t^2 - 1 \\
&= tr\{F^2 S' G_l^{-1} D_0 W^{-1} \frac{1}{T}[\sum_{t=1}^{T} m^l(U_t, h) m^l(U_t, h)' - E m^l(U_t, h) m^l(U_t, h)'] W^{-1} D_0 G_l^{-1} AS\} \\
&\leq |\lambda_{\max}\{\frac{1}{T}[\sum_{t=1}^{T} m^l(U_t, h) m^l(U_t, h)' - E m^l(U_t, h) m^l(U_t, h)']\}| \frac{1}{\lambda_{\min} W} \\
&= O_p(\frac{l}{\sqrt{T}\lambda_{\min} W}) = o_p(1).
\end{aligned}$$

Therefore, we complete the proof that $\sqrt{T} F S' G_l^{-1} D_0 W^{-1} \hat{m}_T(h) \xrightarrow{d} N(0, 1)$. □

Lemma B.0.9. *Suppose $\{Z_i\}$ and $\{Y_i\}$ are two uncorrelated MDS. Suppose for some $\delta > 0$, $\sum_{i=1}^{\infty} i^2 \beta^{\frac{\delta}{1+\delta}}(i) < \infty$, and for all $1 < s, t \leq T$, $j = 1, \cdots, l$, we have 1) $\Psi \equiv E(Z_t Z_t')$, $E(Y_t Y_t') = I$, $T a_T' a_T \to H$, $T^2 tr(\Psi) \to \Lambda$; 2) $E(Z_t' Y_t) = 0$, $T[E(Z_t' Y_t)^{2+\delta/2}]^{\frac{4}{\delta+4}} \to 0$; 3) , $T^2[E|Z_s' Y_1|^{2(1+\delta)}]^{\frac{1}{1+\delta}} \to 0$, $T^2[\lambda_{\max}^{2(1+\delta)}(\Psi) E||y_s||^{4(1+\delta)} + E||Z_s||^{4(1+\delta)}]^{\frac{1}{2(1+\delta)}} \to 0$; 3)$T^{\frac{3}{2}}[E|Z_s' Y_1|^{4(1+\delta)}]^{\frac{1}{2(1+\delta)}}$.*

$$\sum_{t=1}^{T} a_T' Y_t + \sum_{i,j=1}^{T} Z_i' Y_j \to^d N(0, H + \Lambda^*)$$

Proof of Lemma B.0.9 This proof modifies the proofs of Lemma A2 of Newey (2004) and theorem 1 of Hall (1984). Let w denote all possible data for a single observation that includes all of the elements of Y, and $H_T(w_i, w_j) = Z_i'Y_j + Z_j'Y_i$. Then,

$$\sum_{i=1}^{T} a_T'Y_i + \sum_{i,j=1}^{T} Z_i'Y_j = \sum_{i=2}^{T}(A_{iT} + B_{iT}) + R_T,$$

where $A_{iT} = a_T'Y_i$, $B_{iT} = \sum_{j<i} H_T(w_i, w_j)$, and $R_T = \sum_{i=1}^{T} Z_i'Y_i + a_T'Y_1$. First, we have, by $E(Z_i'Y_i) = 0$ and the Davydov inequality,

$$\begin{aligned} E[(\sum_{i=1}^{T} Z_i'Y_i)^2] &\leq TE(Z_i'Y_i)^2 + T\sum_{\tau=1}^{T-1}(1 - \frac{\tau}{T})\alpha(\tau)^{\frac{\delta}{4+\delta}}[E(Z_i'Y_i)^{2+\delta/2}]^{\frac{4}{\delta+4}} \\ &\leq TE(Z_i'Y_i)^2 + T[E(Z_i'Y_i)^{2+\delta/2}]^{\frac{4}{\delta+4}}\sum_{\tau=1}^{\infty}\alpha(\tau)^{\frac{\delta}{4+\delta}} \to 0. \end{aligned}$$

Also, $E(a_T'Y_1)^2 = a_T'EY_1Y_1'a_T \to 0$. By Markov's inequality, we have

$$ER_T \leq E|\sum_{i=1}^{T} Z_iY_i| + E|a_T'Y_1| \to 0.$$

In addition,

$$E(A_{iT})^2 = a_T'E(Y_iY_i')a_T = a_T'a_T,$$

The rest of the proof focuses on two targets, that (a):$\frac{1}{T}\sum_{i=2}^{T}(A_{iT} + B_{iT})^2 \to^p H + \Lambda$ and (b): $s_T^{-2}\sum_{i=2}^{T} E[(A_{iT} + B_{iT})^2 I\{|A_{iT} + B_{iT}| > \epsilon s_T\}] \to 0$ as $T \to \infty$ for every $\epsilon > 0$.

To prove these two results, we first need to establish the asymptotic distribution of $\sum_{i=2} B_{iT}$ as $T \to \infty$. We claim that as $T \to \infty$, $\sum_{i=2}^{T} B_{iT} \xrightarrow{d} N(0, \frac{1}{2}T^2E(Z_iZ_i'))$. We apply the Hjellvik et al. (1998) central limit theorem for degenerate U-statistics, which states that the above statement holds true for some $\delta > 0$, $\sum_{j=1}^{\infty} j^2\beta^{\frac{\delta}{1+\delta}}(j) <$

∞, and

$$\max \frac{1}{\sigma_{BT}^2} T^2 [M_{T1}^{\frac{1}{1+\delta}} + M_{T5}^{\frac{1}{2(1+\delta)}} + M_{T6}^{\frac{1}{2}}] \to 0$$
$$\max \frac{1}{\sigma_{BT}^2} T^{\frac{3}{2}} [M_{T2}^{\frac{1}{2(1+\delta)}} + M_{T3}^{\frac{1}{2}} + M_{T4}^{\frac{1}{2(1+\delta)}}] \to 0,$$

where $\sigma_{BT}^2 = \sum_{1\leq s<t\leq T} \sigma_T^2$, and

$$\begin{aligned}
M_{T1} =& \max\max\nolimits_{1<s<r\leq T}\{E|h(\omega_1,\omega_r)h(\omega_s,\omega_r)|^{1+\delta},\\
&\int |h(\omega_1,\omega_r)h(\omega_s,\omega_r)|^{1+\delta} dP(\omega_1)dP(\omega_s,\omega_r)\},\\
M_{T2} =& \max\max\nolimits_{1<s<r\leq T}\{E|h(\omega_1,\omega_r)h(\omega_s,\omega_r)|^{2(1+\delta)},\\
&\int |h(\omega_1,\omega_r)h(\omega_s,\omega_r)|^{2(1+\delta)} dP(\omega_1)dP(\omega_s,\omega_r),\\
&\int |h(\omega_1,\omega_r)h(\omega_s,\omega_r)|^{2(1+\delta)} dP(\omega_1,\omega_r)dP(\omega_r),\\
&\int |h(\omega_1,\omega_r)h(\omega_s,\omega_r)|^{2(1+\delta)} dP(\omega_1)dP(\omega_s)dP(\omega_r)\},\\
M_{T3} =& \max\max\nolimits_{1<s<t\leq T} E|h(\omega_1,\omega_r)h(\omega_s,\omega_r)|^2,\\
M_{T4} =& \max\max\nolimits_{1<s,r,t\leq T}\{\max_P \int |h(\omega_1,\omega_r)h(\omega_s,\omega_t)|^{2(1+\delta)} dP\},\\
M_{T5} =& \max\max\nolimits_{1<s<r\leq T}\{E|\int h(\omega_1,\omega_s)h(\omega_1,\omega_r)dP(\omega_1)|^{2(1+\delta)},\\
&\int |\int h(\omega_1,\omega_s)h(\omega_1,\omega_r)dP(\omega_1)|^{2(1+\delta)} dP(\omega_s)dP(\omega_r)\},\\
M_{T6} =& \max\max\nolimits_{1<s<t\leq T} E|\int h(\omega_1,\omega_s)h(\omega_1,\omega_r)dP(\omega_1)|^2.
\end{aligned}$$

Let $\sigma_T^2 = E[H^2(\tilde{\omega}_1,\tilde{\omega}_2)]$, and $\{\tilde{\omega}_t\}$ is i.i.d. distributed with the same marginal distribution as $\{\omega_t\}$. It follows that

$$\begin{aligned}
\sigma_T^2 \equiv E[H^2(\tilde{\omega}_1,\tilde{\omega}_2)] &= E[\tilde{Z}_1{}'\tilde{Y}_2\tilde{Y}_2'\tilde{Z}_1 + \tilde{Y}_1'\tilde{Z}_2\tilde{Z}_2'\tilde{Y}_1 + \tilde{Z}_1'\tilde{Y}_2\tilde{Y}_1'\tilde{Z}_2 + \tilde{Z}_2'\tilde{Y}_1\tilde{Y}_2'\tilde{Z}_1]\\
&= 2tr[\tilde{Z}_1\tilde{Z}_1'] = 2tr(\Psi).
\end{aligned}$$

It follows from the proof of Theorem A in Hjellvik et al. (1998) that $\sigma_{BT}^2 = \frac{T^2}{2}\sigma_T^2[1+o(1)] = T^2 tr(\Psi)$. Now, we verify these conditions one by one. First,

$$M_{T1} \leq \max_{1<s\leq T} Eh(\omega_1, \omega_s)^{2(1+\delta)} \leq C[E|Z_1'Y_s|^{2(1+\delta)} + E|Z_s'Y_1|^{2(1+\delta)}] = O(E|Z_s'Y_1|^{2(1+\delta)}).$$

$$M_{T2} \leq \max\max{}_{1<s<r\leq T} E|h(\omega_1, \omega_r)|^{4(1+\delta)} \leq C[E|Z_1'Y_s|^{4(1+\delta)} + E|Z_s'Y_1|^{4(1+\delta)}] = O(E|Z_s'Y_1|^{4(1+\delta)}).$$

Similarly, it is easy to show that $M_{T3} \leq C[E|Z_1'Y_s|^4 + E|Z_s'Y_1|^4] = O(E|Z_s'Y_1|^4)$, and $M_{T4} = O(E|Z_s'Y_1|^{4(1+\delta)})$. For ease of notations, we further define

$$\begin{aligned} G(\omega, \tilde{\tilde{\omega}}) &\equiv E(H_T(\omega_1, \omega)H_T(\omega_1, \tilde{\tilde{\omega}})) = E[(y'Z_1 + z'Y_1)(Z_1'\tilde{\tilde{y}} + Y_1'\tilde{\tilde{z}})] \\ &= y'E(Z_1Z_1')\tilde{\tilde{y}} + z'E(Y_1Y_1')\tilde{\tilde{z}} + y'E(Z_1Y_1')\tilde{\tilde{z}} + z'E(Y_1Z_1')\tilde{\tilde{y}} \\ &= y'\Psi\tilde{y} + z'\tilde{\tilde{z}} \end{aligned}$$

Using the above result, it is immediate that

$$\begin{aligned} M_{T5} &= \max_{1<s<r\leq T}\{E|G(\omega_s, \omega_r)|^{2(1+\delta)}, \int |G(\omega_s, \omega_r)|^{2(1+\delta)} dP(\omega_s)dP(\omega_r)\} \\ &\leq C[E|y_s'\Psi y_t|^{2(1+\delta)} + E|Z_t'Z_s|^{2(1+\delta)}] = O(\lambda_{\max}^{2(1+\delta)}(\Psi)E||y_s||^{4(1+\delta)} + E||Z_s||^{4(1+\delta)}). \end{aligned}$$

Lastly, we have

$$M_{T6} = \max_{1<s<r\leq T} EG^2(\omega_s, \omega_r) = O(\lambda_{\max}^2(\Psi)E||Y_t||^4 + E||Z_t||^4).$$

By assumptions 1)-3), we have

$$\begin{aligned} &T^2[M_{T1}^{\frac{1}{1+\delta}} + M_{T5}^{\frac{1}{2(1+\delta)}} + M_{T6}^{\frac{1}{2}}] \to 0 \\ &T^{\frac{3}{2}}[M_{T2}^{\frac{1}{2(1+\delta)}} + M_{T3}^{\frac{1}{2}} + M_{T4}^{\frac{1}{2(1+\delta)}}] \to 0, \end{aligned}$$

Therefore, all the assumptions in Theorem 1 by Gao and Hong (2007) are satisfied. Thus we proved that

$$\sum_{i=2}^{T} B_{iT} \xrightarrow{d} N(0, \frac{1}{2}T^2\sigma_T^2).$$

Let $V_{BT} = \frac{1}{2}T^2\sigma_T^2 = T^2 tr(\Psi) \equiv \Lambda$. It implies two important results that we are going to further take advantage of in the next stage. Namely, First

$$V_{BT}^{-1} E(\sum_{i=2}^{T} B_{iT})^2 \xrightarrow{p} 1,$$

and Second,

$$V_{BT}^{-2} \sum_{i=2}^{T} E[B_{iT}^2 I\{|B_{iT}| > \varepsilon V_{BT}\}] \xrightarrow{p} 0.$$

Lastly, we want to show that 1) $s_T^{-2} \sum_{i=2}^{T} E[(A_{iT} + B_{iT})^2 I\{|A_{iT} + B_{iT}| > \epsilon s_T\}] \to 0$ as $T \to \infty$ for every $\varepsilon > 0$, and 2)$\frac{1}{T} \sum_{i=1}^{T} (A_{iT} + B_{iT})^2 \xrightarrow{p} H + \Lambda$. Note that $TE|a_{iT}' Y_i|^2 \to 0$, and it is immediately that

$$\begin{aligned}
&s_T^{-2} \sum_{i=2}^{T} E[(A_{iT} + B_{iT})^2 I\{|A_{iT} + B_{iT}| > \epsilon s_T\}] \\
&\leq CH^{-2} \sum_{i=2}^{T} E(A_{iT})^2 I\{|A_{iT}| > \epsilon s_T/2\} + CV_{BT}^{-2} \sum_{i=2}^{T} E(B_{iT})^2 I\{|B_{iT}| > \epsilon s_T/2\}] \to 0.
\end{aligned}$$

To establish part (2), we note that $||\frac{1}{T} \sum_{i=1}^{T} Y_i Y_i' - EY_i Y_i'|| = o_p(1)$, thus

$$||\frac{1}{T} \sum_{i=2}^{T} A_{iT} - H|| = o_p(1).$$

$$\frac{1}{T} \sum_{i=1}^{T} (A_{iT} + B_{iT})^2 - (H + \Lambda) = \frac{1}{T} \sum_{t=2}^{T} B_{iT}^2 - \Lambda + \frac{2}{T} \sum_{t=2}^{T} A_{iT} B_{iT} \xrightarrow{p} 0.$$

By Brown (1971) martingale central limit theorem, the conclusion follows directly.

□

Lemma B.0.10. *Suppose Assumptions 3.3.1-3.3.4 hold, and $\hat{h}_l$ is a two-stage efficient series estimator with W being a nonstochastic weighting function. Then for any given $x \in \mathbf{X}$, as $l \to \infty$ with $T \to \infty$,*

$$\frac{\partial \hat{Q}_T(h_{l0})}{\partial \gamma^l} \frac{\lambda_{\min}(W)T}{l} \xrightarrow{d} N(0, H).$$

Proof of Lemma B.0.10 We take first order conditions for the two-stage efficient GMM series estimator $\hat{h}_l$. For some $c \in R^l$, we have

$$0 = \frac{\partial \hat{Q}_T(\hat{h}_l)}{\partial \gamma_j}, \quad j = 1, \cdots, l;$$
$$c' \frac{\partial \hat{Q}_T(h_l)}{\gamma_j} c \frac{\lambda_{\min}(W)T^2}{l} = c' \frac{\partial \hat{m}_T(h_l)'}{\partial \gamma_j} \hat{W}^{-1}(h_l) \hat{m}_T(h_l) c \frac{\lambda_{\min}(W)T^2}{l}.$$

where and $a_T' = c'\hat{D}^l(h_{lo})'\hat{W}^{-\frac{1}{2}}(h_{lo})\sqrt{\frac{\lambda_{\min}(W)}{T}}$, and $Y_{iT} = \sqrt{\lambda_{\min}(W)}\hat{W}^{-\frac{1}{2}}(h_{lo})m(U_i, h_{lo})$. Let $\hat{D}^l(h_{lo}) = [\hat{D}_1(h_{lo}), \cdots, \hat{D}_l(h_{lo})]$ where $\hat{D}_j(h_{lo}) = \frac{\sqrt{T}}{l}[\frac{\partial \hat{m}_T(h_{lo})}{\partial \gamma_j}]$, and $D^l(h_{lo}) = [D_1(h_{lo}), \cdots, D_l(h_{lo})]$ where $D_j(h_l) = E[\hat{D}_j(h_l)]$. To establish the central limit theorem, we need to verify the following two conditions by Brown (1971), namely

$$H^{-2} \sum_{t=1}^{T} E(a_T' Y_{iT})^2 I\{|a_T' Y_{iT}| > \epsilon H\} \xrightarrow{p} 0,$$

and

$$H^{-1} E[\sum_{t=1}^{T} a_T' Y_{iT}]^2 \xrightarrow{P} 1,$$

where $H = Ta_T'a_T$ has $||H|| = O(1)$ as proved.

First, because $m(U_i, h)$ is a martingale difference sequence, $EY_{iT}Y_{iT}' = I$, an $l \times l$ identity matrix. Therefore the second requirement holds automatically, that

$$H^{-1}[\sum_{t=1}^{T} a_T' Y_{iT}]^2 = H^{-1}Ta_T'E[Y_{iT}Y_{iT}'] + 2H^{-1}\sum_{s<t} a_T'E(Y_{sT}Y_{tT}')a_T$$
$$= H^{-1}Ta_T'a_T = 1.$$

Then we focus on the proof of the Lindeberg's condition. By the c_r inequality and Cauchy-Schwarz inequality, we have

$$\begin{aligned}
&H^{-2}\sum_{t=1}^{T} E(a_T'Y_{iT})^2 I\{|a_T'Y_{iT}| > \epsilon H\}\\
&= H^{-2}(\epsilon H)^{-\delta/2}\sum_{t=1}^{T} E(a_T'Y_{iT})^{2+\delta} \leq CTc'E|\sum_{j=1}^{l}\hat{D}_j(h_{lo})'m_j(U_t,h_{lo})/T|^{2+\delta}c\\
&\leq (\frac{1}{T})^{1+\delta} c'c\sum_{j=1}^{l}\sum_{k=1}^{l}[E|\frac{\partial m_j(U_t,h_l)}{\partial\gamma_k}|^{4+2\delta}]^{\frac{1}{2}}[E|m_j(U_t,h_l)|^{4+2\delta}]^{\frac{1}{2}} = o_p(1).
\end{aligned}$$

Therefore, we complete the proof. □

Lemma B.0.11. *Suppose Assumptions 3.3.1-3.3.6 hold, and $\hat{h}_l$ is a CUE GMM series estimator. Then for any given $x \in \mathbf{X}$, as $l \to \infty$ with $T \to \infty$,*

$$\frac{\partial\hat{Q}_T(h_{l0})}{\partial\gamma^l}\frac{\lambda_{\min}(W)T}{l} \xrightarrow{d} N(0, H+\Lambda).$$

Proof of Lemma B.0.10 We take first order conditions for the CUE GMM series estimator $\hat{h}_l$, which follows that

$$\begin{aligned}
&0 = \frac{\partial\hat{Q}_T(\hat{h}_l)}{\partial\gamma_j}, \quad j = 1,\cdots,l;\\
&\frac{\partial\hat{Q}_T(h_l)}{\gamma_j} = \{\frac{\partial\hat{m}_T(h_l)'}{\partial\gamma_j}\hat{W}^{-1}(h_l)\hat{m}_T(h_l) - \hat{m}_T(h_l)'\hat{W}^{-1}(h_l)\frac{\partial\hat{W}(h_l)}{\partial\gamma_j}\hat{W}^{-1}(h_l)\hat{m}_T(h_l)\}.
\end{aligned}$$

Without abuse of notations, we denote for $j = 1,\cdots,l$ and $t = 1,\cdots,T$,

$$\begin{aligned}
\hat{A}_j(h_l) &= \frac{\partial\hat{W}(h_l)}{\partial\gamma_j}\hat{W}^{-1}(h_l),\\
A_j(h_{lo}) &= \frac{\partial W(h_{lo})}{\partial\gamma_j}W^{-1}(h_{lo}) = Tcov[\frac{\partial\hat{m}_T(h_{lo})}{\partial\gamma_j}, \hat{m}_T(h_{lo})]W^{-1},\\
\bar{U}^j(h_l) &= \sqrt{T}\{\frac{\partial\hat{m}_T(h_{lo})}{\partial\gamma_j} - E[\frac{\partial\hat{m}_T(h_{lo})}{\partial\gamma_j}] - \hat{A}_j\hat{m}_T(h_{lo})\}\\
U_t^j(h_{lo}) &= \{\frac{\partial m(U_t,h_{lo})}{\partial\gamma_j} - E[\frac{\partial m(U_t,h_{lo})}{\partial\gamma_j}] - A_j m(U_t,h_{lo})\}.
\end{aligned}$$

Let $\hat{D}^l(h_{lo}) = [\hat{D}_1(h_{lo}), \cdots, \hat{D}_l(h_{lo})]$ where $\hat{D}_j(h_{lo}) = \frac{\sqrt{T}}{l}[\frac{\partial \hat{m}_T(h_{lo})}{\partial \gamma_j}]$, $D^l(h_{lo}) = [D_1(h_{lo}), \cdots, D_l(h_{lo})]$ where $D_j(h_l) = E[\hat{D}_j(h_l)]$, and $\hat{H}(h_l) = \frac{T}{l^2}\frac{\partial^2 \hat{Q}_T(h_l)}{\partial \gamma^l \partial \gamma^{l\prime}}$. Then, for all $j = 1, \cdots, l$, we have

$$
\begin{aligned}
& l\frac{\partial \hat{Q}_T(h_{lo})}{\partial \gamma_j}\frac{\lambda_{\min}(W)T}{l^2} \\
&= \{\frac{\partial \hat{m}_T(h_{lo})'}{\partial \gamma_j}\hat{W}^{-1}(h_{lo})\hat{m}_T(h_{lo}) - \hat{m}_T(h_{lo})\hat{A}_j'(h_{lo})\hat{W}^{-1}(h_{lo})\hat{m}_T(h_{lo})\}\frac{\lambda_{\min}(W)T}{l} + o_p(1) \\
&= \hat{D}_j(h_{lo})'\hat{W}^{-1}(h_{lo})\lambda_{\min}(W)\sqrt{T}\hat{m}_T(h_{lo}) + \bar{U}^{j\prime}\hat{W}^{-1}(h_{lo})\lambda_{\min}(W)\sqrt{T}\hat{m}_T(h_{lo})/l + o_p(1).
\end{aligned}
$$

Let $c \in R^l$, and $a_T' = c'\hat{D}^l(h_{lo})'\hat{W}^{-\frac{1}{2}}((h_{lo}))\sqrt{\frac{\lambda_{\min}(W)}{T}}$, $Y_{iT} = \sqrt{\lambda_{\min}(W)}\hat{W}^{-\frac{1}{2}}(h_{lo})m(U_i, h_{lo})$, and $Z_{iT} = \sqrt{\lambda_{\min}(W)}\hat{W}^{-\frac{1}{2}}(h_{lo})\bar{U}_i(h_{lo})c\frac{1}{Tl}$. Therefore, the first order condition can be summarized as follows:

$$
lc'\frac{\partial \hat{Q}_T(h_{lo})}{\partial \gamma^l}c\frac{\lambda_{\min}(W)T}{l^2} = [\sum_{i=1}^{T} a_T'Y_{iT} + \sum_{i,j=1}^{T} Z_{iT}'Y_{jT}] + o_p(1).
$$

By Lemma B.0.7 established above, we first have

$$
Ta_T'a_T = c'D_T(h_l)'W^{-1}\lambda_{\min}(W)D_T(h_l)c \to c'Hc = O_p(1),
$$

in addition, letting $\Lambda = E[U_i'W^{-1}(h_{lo})U_i]\lambda_{\min}(W)$, we have:

$$
T^2tr(\Psi) = T^2trE(Z_{iT}'Z_{iT}) = T^2\frac{1}{T^2l^2}c'E(U_i'W^{-1}(h_{lo})U_i)\lambda_{\min}(W) = c'\Lambda c.
$$

Note that by assuming for all $1 \leq j \leq l$ and $1 \leq t \leq T$, $\lambda_{\max}E[\frac{\partial m^l(U_t,h_l)}{\partial \gamma_j}\frac{\partial m^l(U_t,h_l)}{\partial \gamma_j'}]$ is uniformly bounded by a constant smaller than infinity, we have

$$
c'\Lambda c \leq \frac{C}{l^2}||E(U_i'U_i)|| \leq \frac{C}{l^2}\sum_{i=1}^{l}\sum_{j=1}^{l}E[\frac{\partial m^l(U_t,h_l)}{\partial \gamma_j}\frac{\partial m^l(U_t,h_l)}{\partial \gamma_i}] = O_p(1),
$$

$$
\lambda_{\max}(\Psi) =\leq \frac{C}{T^2l^2}\lambda_{\max}E(U_i'U_i) \leq \frac{C}{T^2l^2}\sum_{j=1}^{l}\lambda_{\max}E[\frac{\partial m^l(U_t,h_l)}{\partial \gamma_j}\frac{\partial m^l(U_t,h_l)}{\partial \gamma_j'}] = O_p(\frac{1}{T^2l}).
$$

Now, to establish the conclusion, we start validating all the assumptions in Lemma B.0.9. Using the assumption that $E|\frac{m_j(U_t,h_l)}{\gamma_k}|^{4(1+\delta)}$ and $E|m_j(U_t,h_l)|^{4(1+\delta)}$ are uniformly bounded by a constant smaller than infinity for $1 \leq j,k \leq l$, it is immediate that from the c_r inequality in White (1999) and the Markov's inequality,

$$E|\frac{\partial m^l(U_t,h_l)}{\partial \gamma_k'} m^l(U_t,h_l)|^{2(1+\delta)} \leq \sup_{1\leq j,k\leq l} lE[\frac{\partial m_j^{4(1+\delta)}(U_t,h_l)}{\partial \gamma_k}]^{\frac{1}{2}} E[m_j^{4(1+\delta)}(U_t,h_l)]^{\frac{1}{2}} = O(l),$$

Applying the c_r inequality in White (1999) again, for some $0 < \delta < 1$, any $1 \leq t, s \leq T$, we have:

$$T^2\{E|Z_{tT}'Y_{sT}|^{2(1+\delta)}\}^{\frac{1}{1+\delta}} \leq T^2\{\frac{1}{Tl}E|c'U_i m^l(U_t,h_l)|^{2(1+\delta)}\}^{\frac{1}{1+\delta}}$$
$$= \frac{1}{l^2}\{E|\sum_{k=1}^{l} c_j[\frac{\partial m^l(U_t,h_l)}{\partial \gamma_k'} m^l(U_t,h_l)]|^{2(1+\delta)}\}^{\frac{1}{1+\delta}} \leq \frac{c_r}{l^2}\{\sum_{k=1}^{l} E[\frac{\partial m^l(U_t,h_l)}{\partial \gamma_k'} m^l(U_t,h_l)]|^{2(1+\delta)}\}^{\frac{1}{1+\delta}}$$
$$= Cl^{-2}l^{\frac{2}{1+\delta}} \to 0$$

When $E|\frac{m_j(U_t,h_l)}{\gamma_k}|^{8(1+\delta)}$ and $E|m_j(U_t,h_l)|^{8(1+\delta)}$ are uniformly bounded by a constant smaller than infinity for all $1 \leq j,k \leq l$, we can show that $T^{\frac{3}{2}}[E|Z_s'Y_1|^{4(1+\delta)}]^{\frac{1}{2(1+\delta)}} \to 0$ using the similar argument. Analogously, we can also show that

$$E|[\frac{\partial m^l(U_t,h_l)}{\gamma_j'}][\frac{\partial m^l(U_t,h_l)}{\gamma_j}]|^{2(1+\delta)} \leq l \sup_{1\leq k\leq l} E|\frac{\partial m_k(U_t,h_l)}{\gamma_j}|^{2(1+\delta)} = O(l).$$

Then, we verify the last condition, that for all $1 < s \leq T$, $T^2[\lambda_{\max}^{2(1+\delta)}(\Psi)E|Y_t'Y_s|^{2(1+\delta)} + E|Z_t'Z_s|^{2(1+\delta)}]^{\frac{1}{2(1+\delta)}} \to 0$. Applying the c_r inequality here again, for all $1 \leq t \neq s \leq l$, we have

$$
\begin{aligned}
&T^2[\lambda_{\max}^{2(1+\delta)}(\Psi)E|Y_t'Y_s|^{2(1+\delta)} + E|Z_t'Z_s|^{2(1+\delta)}]^{\frac{1}{2(1+\delta)}} \\
&\leq c_r T^2[\lambda_{\max}^{2(1+\delta)}(\Psi)E|Y_t'Y_s|^{2(1+\delta)}]^{\frac{1}{2(1+\delta)}} + T^2[E|Z_t'Z_s|^{2(1+\delta)}]^{\frac{1}{2(1+\delta)}} \\
&\leq T^2\frac{1}{T^2 l}[E|Y_t'Y_s|^{2(1+\delta)}]^{\frac{1}{2(1+\delta)}} + T^2\frac{1}{T^2 l^2}[E|U_i'U_i|^{2(1+\delta)}]^{\frac{1}{2(1+\delta)}} \\
&\leq \frac{1}{l}[E|\sum_{k=1}^{l} m_k(U_t, h_l)m_k(U_s, h_l)|^{2(1+\delta)}]^{\frac{1}{2(1+\delta)}} + \frac{1}{l^2}\{E|\sum_{j=1}^{l}[\frac{\partial m^l(U_t, h_l)}{\gamma_j'}][\frac{\partial m^l(U_t, h_l)}{\gamma_j}]|^{2(1+\delta)}\}^{\frac{1}{2(1+\delta)}} \\
&\leq \frac{c_r}{l}[\sum_{k=1}^{l} E|m_k(U_t, h_l)m_k(U_s, h_l)|^{2(1+\delta)}]^{\frac{1}{2(1+\delta)}} + \frac{c_r}{l^2}\{\sum_{j=1}^{l} E|[\frac{\partial m^l(U_t, h_l)}{\gamma_j'}][\frac{\partial m^l(U_t, h_l)}{\gamma_j}]|^{2(1+\delta)}\}^{\frac{1}{2(1+\delta)}} \\
&= l^{-[1-\frac{1}{2(1+\delta)}]} + l^{-[2-\frac{l^2}{2(1+\delta)}]} \to 0
\end{aligned}
$$

Thus all the assumptions in Lemma B.0.9 are satisfied, we reach the conclusion that

$$
l\frac{\partial \hat{Q}_T(h_{lo})}{\partial \gamma_j}\frac{\lambda_{\min}(W)T}{l^2} \xrightarrow{d} N(0, H + \Lambda)
$$

□

Proof of Theorem 3.3.5 Suppose Assumptions hold and suppose $\frac{l^4}{\lambda_{\min}^6(W)T} \to 0$. Let $H \equiv D(h^o)'W^{-1}\lambda_{\min}(W)D(h^o)$. Then we need to show that for any $\hat{h}_l \xrightarrow{p} h^o$, $\frac{\partial^2 \hat{Q}_T(h_l)}{\partial \gamma_l \partial \gamma_l'}\frac{\lambda_{\min}(W)T}{l^2} \xrightarrow{p} H$, as T goes to infinity. We modify the proof in Lemma A12 by Newey (2004) to make it compatible with the framework in this paper as follows. First, for notational convenience, we drop γ^l's l argument, and let k and j denote the derivatives with respect to γ_k^l and γ_j^l for $1 \leq k, j \leq l$ as

$$\tilde{m} \equiv \tilde{m}^l(h_l) = \frac{\sqrt{T}}{l}\hat{m}_T(h_l), \quad \tilde{m}_{(k)} \equiv \tilde{m}^l_{(k)}(h_l) = \frac{\partial \tilde{m}^l(h_l)}{\gamma_k}, \quad \tilde{m}_{(k,j)} \equiv \tilde{m}^l_{(k,j)} = \frac{\partial^2 \tilde{m}(h_l)}{\gamma_k \gamma_j}$$

$$\hat{W}_{(k)} \equiv \frac{\partial \hat{W}(h_l)}{\partial \gamma_k}, \quad \hat{W}_{(k,j)} \equiv \frac{\partial^2 \hat{W}(h_l)}{\partial \gamma_k \partial \gamma_j}$$

$$\hat{Q}_{(k)} \equiv \frac{\partial \hat{Q}_T(h_l, W(h_l))}{\partial \gamma_k} = \tilde{m}'_{(k)}\hat{W}^{-1}(h_l)\lambda_{\min}(W)\tilde{m} - \frac{1}{2}\tilde{m}'\hat{W}^{-1}(h_l)\lambda_{\min}(W)\hat{W}_{(k)}\hat{W}^{-1}(h_l)\tilde{m},$$

$$\begin{aligned}\hat{Q}_{(k,j)} \equiv \frac{\partial^2 \hat{Q}(h_l, W(h_l))}{\partial \gamma_{lk}\partial \gamma_j} &= \tilde{m}'_{(k,j)}\hat{W}^{-1}(h_l)\lambda_{\min}(W)\tilde{m} + \tilde{m}'_{(k)}\hat{W}^{-1}(h_l)\lambda_{\min}(W)\tilde{m}_{(j)} \\ &- \tilde{m}'_{(k)}\hat{W}^{-1}(h_l)\hat{W}_{(j)}\hat{W}^{-1}(h_l)\lambda_{\min}(W)\tilde{m} - \tilde{m}'_{(j)}\hat{W}^{-1}(h_l)\lambda_{\min}(W)\hat{W}_{(k)}\hat{W}^{-1}(h_l)\tilde{m} \\ &+ \tilde{m}'\hat{W}^{-1}(h_l)\hat{W}_{(j)}\hat{W}^{-1}(h_l)\lambda_{\min}(W)\hat{W}_{(k)}\hat{W}^{-1}(h_l)\tilde{m} - \tilde{m}'\hat{W}^{-1}(h_l)\lambda_{\min}(W)\hat{W}_{(k,j)}\hat{W}^{-1}(h_l)\tilde{m}.\end{aligned}$$

Note that for $\tilde{Q} \equiv \frac{1}{2}\tilde{m}'W^{-1}\tilde{m}$, and $\hat{Q}_{(k,j)} \equiv \frac{\partial^2 \hat{Q}(\gamma_l)}{\partial \gamma_k \partial \gamma_j}$ has the same formula as $\hat{Q}_{(k,j)}$ with $W = W(h_l)$. We want to show that $||\hat{Q}(\hat{h}_l, W(h_l))_{\gamma^l\gamma^{l\prime}} - \tilde{Q}(h_{lo}, W(h_{lo}))_{\gamma^l\gamma^{l\prime}}|| \xrightarrow{p} 0$. By the Triangular inequality,

$$\begin{aligned}&||\hat{Q}(\hat{h}_l, W(h_l))_{\gamma^l\gamma^{l\prime}} - \tilde{Q}(h_{lo}, W(h_{lo}))_{\gamma^l\gamma^{l\prime}}|| \\ &\leq ||\hat{Q}(\hat{h}_l, W(h_l))_{\gamma^l\gamma^{l\prime}} - \tilde{Q}(\hat{h}_l, \hat{W}(h_l))_{\gamma^l\gamma^{l\prime}}|| + ||\tilde{Q}(\hat{h}_l, \hat{W}(h_l))_{\gamma^l\gamma^{l\prime}} - \tilde{Q}(h_{lo}, W(h_{lo}))_{\gamma^l\gamma^{l\prime}}||.\end{aligned}$$

Because $\tilde{Q}(\gamma_l)_{\gamma_l\gamma_l'}$ is assumed stochastic equicontinuous, it is sufficient to prove $||\hat{Q}(\hat{\gamma}_l)_{\gamma_l\gamma_l'} - \tilde{Q}(\hat{\gamma}_l)_{\gamma_l\gamma_l'}|| \xrightarrow{p} 0$. It follows that for each pair of (k, j) with $1 \leq k, j \leq l$, by Lemma B.0.4,

$$\begin{aligned}&\sup_{\gamma_l} ||\tilde{m}'_{(k,j)}\hat{W}^{-1}\lambda_{\min}(W)\tilde{m} - \tilde{m}'_{(k,j)}W^{-1}\lambda_{\min}(W)\tilde{m}|| \\ &\leq \sup_{\gamma_l} \lambda_{\min}W||\tilde{m}_{(k,j)}|| \sup_{\gamma_l} ||\hat{W}^{-1} - W^{-1}|| \sup_{\gamma_l} ||\tilde{m}|| \\ &= O_p(1)\lambda_{\min}(W) \sup_{\gamma_l} ||W^{-1}(\hat{W} - W)\hat{W}^{-1}||O_p(1) \\ &\leq O_p(\frac{l}{\lambda_{\min}(W)\sqrt{T}}) = o_p(1).\end{aligned}$$

In addition, we have

$$\begin{aligned}
&\sup_{\gamma_l} ||\tilde{m}'_{(k)}\hat{W}^{-1}\lambda_{\min}(W)\tilde{m}_{(j)} - \tilde{m}'_{(k)}W^{-1}\lambda_{\min}(W)\tilde{m}_{(j)}|| \\
&\leq \lambda_{\min}(W)\sup_{\gamma_l}||\tilde{m}_{(k)}||\sup_{\gamma_l}||W^{-1}(\hat{W}-W)\hat{W}^{-1}||\sup_{\gamma_l}||\tilde{m}_{(j)}|| \\
&= O_p(\frac{l}{\lambda_{\min}(W)}\sqrt{T}) = o_p(1).
\end{aligned}$$

By Triangular inequality, it follows that

$$\begin{aligned}
&\sup_{\gamma_l} ||\tilde{m}'_{(k)}\hat{W}^{-1}\hat{W}_{(j)}\hat{W}^{-1}\lambda_{\min}(W)\tilde{m} - \tilde{m}'_{(k)}W^{-1}W_{(j)}W^{-1}\lambda_{\min}(W)\tilde{m}|| \\
&\leq \lambda_{\min}(W)\sup_{\gamma_l}||\tilde{m}_{(k)}||\sup_{\gamma_l}||\hat{W}^{-1}\hat{W}_{(j)}\hat{W}^{-1} - W^{-1}W_{(j)}W^{-1}||\sup_{\gamma_l}||\tilde{m}|| \\
&\leq O_p(1)\lambda_{\min}(W)\{||[\hat{W}^{-1}-W^{-1}]\hat{W}_{(k)}\hat{W}^{-1}|| + ||W^{-1}[\hat{W}_{(k)} - W_{(k)}]\hat{W}^{-1}|| \\
&+ ||W^{-1}W_{(k)}[\hat{W}^{-1}-W^{-1}]||\} \\
&= O_p(\frac{l}{\lambda^2_{\min}(W)\sqrt{T}}).
\end{aligned}$$

Similarly, we can show that

$$\begin{aligned}
&\sup_{\gamma_l}\lambda_{\min}(W)||\tilde{m}'_{(j)}\hat{W}^{-1}\hat{W}_{(k)}\hat{W}^{-1}\tilde{m} - \tilde{m}'_{(j)}W^{-1}W_{(k)}W^{-1}\tilde{m}|| = O_p(\frac{l}{\lambda^2_{\min}(W)\sqrt{T}}), \\
&\sup_{\gamma_l}\lambda_{\min}(W)||\tilde{m}'\hat{W}^{-1}\hat{W}_{(k,j)}\hat{W}^{-1}\tilde{m} - \tilde{m}'\hat{W}^{-1}\hat{W}_{(k,j)}\hat{W}^{-1}\tilde{m}|| = O_p(\frac{l}{\lambda^2_{\min}(W)\sqrt{T}}).
\end{aligned}$$

We apply the Triangular inequality on the last term, and because $\lambda_{\max}W$ is assumed to be uniformly bounded by some constant smaller than infinity, we have

$$\begin{aligned}
&\sup_{\gamma_l}\lambda_{\min}(W)||\tilde{m}'\hat{W}^{-1}\hat{W}_{(j)}\hat{W}^{-1}\hat{W}_{(k)}\hat{W}^{-1}\tilde{m} - \tilde{m}'\hat{W}^{-1}\hat{W}_{(j)}\hat{W}^{-1}\hat{W}_{(k)}\hat{W}^{-1}\tilde{m}|| \\
&\leq \lambda_{\min}(W)\sup_{\gamma_l}||\tilde{m}||\hat{W}^{-1}\hat{W}^{-1}\hat{W}^{-1} - W^{-1}W^{-1}W^{-1}||\sup_{\gamma_l}||\tilde{m}|| \\
&\leq O_p(1)\{||W^{-1}[\hat{W}-W]\hat{W}^{-1}\hat{W}^{-1}\hat{W}^{-1}|| + ||W^{-1}W^{-1}[\hat{W}-W]\hat{W}^{-1}\hat{W}^{-1}|| \\
&+ ||W^{-1}W^{-1}W^{-1}||\hat{W}-W||\hat{W}^{-1}\} = O_p(\frac{l}{\lambda^3_{\min}(W)\sqrt{T}}).
\end{aligned}$$

Thus, we can show that

$$E||\hat{Q}(\hat{\gamma}_l)_{\gamma_l\gamma_l'} - \tilde{Q}(\hat{\gamma}_l)_{\gamma_l\gamma_l'}||^2 \leq \sum_{k=1}^{l}\sum_{j=1}^{l}|\hat{Q}_{(k,j)} - \tilde{Q}_{(k,j)}||^2$$
$$\leq O_p(\frac{l^4}{\lambda_{\min}^6(W)T}) \xrightarrow{p} 0.$$

Thus, by the Markov's inequality, we have $||\hat{Q}(\hat{\gamma}_l)_{\gamma_l\gamma_l'} - \tilde{Q}(\hat{\gamma}_l)_{\gamma_l\gamma_l'}|| \xrightarrow{p} 0$.

Next, we want to show that $\tilde{Q}(\hat{\gamma}_l)_{\gamma_l\gamma_l'} \xrightarrow{p} H(h_{lo})$. For ease of notations, we drop γ_l and t for convenience. For all pair of $1 \leq k, j \leq l$, denote

$\Upsilon_{(k)} \equiv E[m_{(k)}m'] = E[m_{(k)}(U_t, h_l)m(U_t, h_l)']$, $\quad \Upsilon_{(kj)} \equiv E[m_{(k,j)}(U_t, h_l) - \bar{m}_{(k,j)}(U_t, h_l)]m']$

$\Upsilon_{(k,j)} \equiv E[m_{(k)}(U_t, h_l) - \bar{m}_{(k)}(U_t, h_l)][m_{(j)}(U_t, h_l) - \bar{m}_{(j)}(U_t, h_l)']$.

And note that

$$W_{(k)} = \Upsilon_{(k)} + \Upsilon_{(k)}', \quad W_{(k,j)} = \Upsilon_{(kj)} + \Upsilon_{(k,j)} + \Upsilon_{(kj)}' + \Upsilon_{(k,j)}'.$$

So, we have

$$\tilde{m}_{(k,j)}'W^{-1}\tilde{m} = tr(W^{-1}\Upsilon_{(kj)}') + O_p(\frac{l}{\lambda_{\min}(W)\sqrt{T}}),$$
$$\tilde{m}_{(k)}'W^{-1}\tilde{m}_j = \bar{m}_{(k)}'W^{-1}\bar{m}_{(j)} + tr(W^{-1}\Upsilon_{(k,j)}) + O_p(\frac{l}{\lambda_{\min}(W)\sqrt{T}}),$$
$$\tilde{m}_{(l)}'W^{-1}W_{(k)}W^{-1}\tilde{m} = tr(W^{-1}W_{(k)}W^{-1}\Upsilon_{(j)}') + O_p(\frac{l}{\lambda_{\min}^2(W)\sqrt{T}}),$$
$$\tilde{m}'W^{-1}W_{(k,j)}W^{-1}\tilde{m} = tr(W^{-1}W_{(k,j)}) + O_p(\frac{l}{\sqrt{T}}),$$
$$\tilde{m}'W^{-1}W_jW^{-1}W_jW^{-1}\tilde{m} = tr(W^{-1}W_{(j)}W^{-1}W_{(k)}) + O_p(\frac{l}{\sqrt{T}}).$$

Therefore, plugging these terms into $\tilde{Q}_{(k,j)}$, it yields for each pair of (k, j),

$$
\begin{aligned}
\tilde{Q}_{(k,j)} &= tr(W^{-1}\Upsilon'_{(kj)}) + \bar{m}'_{(k)}W^{-1}\bar{m}_{(j)} + tr(W^{-1}\Upsilon_{(k,j)}) - tr(W^{-1}[\Upsilon_{(j)} + \Upsilon'_{(j)}]W^{-1}\Upsilon'_{(k)}) \\
&- tr(W^{-1}[\Upsilon_{(k)} + \Upsilon'_{(k)}]W^{-1}\Upsilon'_{(j)}) + tr(W^{-1}(\Upsilon_{(k)} + \Upsilon'_{(k)})W^{-1}(\Upsilon_{(j)} + \Upsilon'_{(j)})) \\
&- \frac{1}{2}tr(W^{-1}(\Upsilon_{(kj)} + \Upsilon_{(k,j)} + \Upsilon'_{(k,j)} + \Upsilon'_{(kj)})) + O_p(\frac{l}{\lambda^2_{\min}(W)\sqrt{T}}) \\
&= \bar{m}'_{(k)}W^{-1}\bar{m}_{(j)} + O_p(\frac{l}{\lambda^2_{\min}(W)\sqrt{T}}).
\end{aligned}
$$

Therefore, we have

$$
\begin{aligned}
||\tilde{Q}(\hat{\gamma}_l)_{\gamma_l\gamma_l'} - H(h_{lo})||^2 &\leq \sum_{k=1}^{l}\sum_{j=1}^{l} ||\tilde{Q}_{(k,j)} - H_{(k,j)(h_{lo})}||^2 \\
&\leq O_p(\frac{l^4}{\lambda^4_{\min}(W)T}) = o_p(1).
\end{aligned}
$$

Because we can show that $H(h)$ is stochastic equicontinuous, so by $h_{lo} \to h^o$, by Triangular inequality, we have

$$
||\tilde{Q}(\hat{\gamma}_l) - H(h^o)|| \leq ||||\tilde{Q}(\hat{\gamma}_l) - H(h_{lo})|| + ||H(h_{lo}) - H(h^o)|| \to 0.
$$

□

Proof of Theorem 3.3.5 Using the conclusion from Lemma B.0.10 and Theorem 3.3.4, for any given $x \in \mathbf{X}$, we have

$$
\begin{aligned}
& l(\hat{h}_l(x) - h_{l0}(x)) \\
&= -[\frac{\partial^2\hat{Q}_T(\bar{h}_l)}{\partial\gamma\partial\gamma'}\lambda_{\min}(W)\frac{T}{l^2}]^{-1} l \frac{\partial\hat{Q}_T(h_{lo})}{\partial\gamma}\lambda_{\min}(W)\frac{T}{l^2} \\
&= -\hat{H}(\bar{h}_l)^{-1}[D_j(h_{lo})'W^{-1}\lambda_{\min}(W)\sqrt{T}\hat{m}_T(h_{lo})] \\
&\xrightarrow{d} N(0, H^{-1}).
\end{aligned}
$$

In addition, given Assumption 3.2.2, for all $s > 1$, we have

$$l\big(h_{l0}(x) - h^o(x)\big) = O(l\kappa^{-s}) = o(1).$$

Therefore, we show that $l(\hat{h}_l(x) - h_{l0}(x)) \to N(0, H^{-1})$. $\square$

Proof of Theorem 3.3.6 Using the conclusion from Lemma B.0.11 and Theorem 3.3.4, for any given $x \in \mathbf{X}$, we have

$$\begin{aligned}
&l(\hat{h}_l(x) - h_{l0}(x)) \\
&= -[\frac{\partial^2 \hat{Q}_T(\bar{h}_l)}{\partial\gamma\partial\gamma'}\lambda_{\min}(W)\frac{T}{l^2}]^{-1} l \frac{\partial \hat{Q}_T(h_{lo})}{\partial\gamma}\lambda_{\min}(W)\frac{T}{l^2} \\
&= -H(\bar{h}_l)^{-1}[D_j(h_{lo})'W^{-1}\lambda_{\min}(W)\sqrt{T}\hat{m}_T(h_{lo}) + \bar{U}'W^{-1}\lambda_{\min}(W)\sqrt{T}\hat{m}_T(h_{lo})/l] + o_p(1) \\
&\xrightarrow{d} N(0, H^{-1} + H^{-1}\Lambda^\star H^{-1}).
\end{aligned}$$

In addition, given Assumption 3.2.2, for all $s > 1$, we have

$$l\big(h_{l0}(x) - h^o(x)\big) = O(l\kappa^{-s}) = o(1).$$

Therefore, we show that $l(\hat{h}_l(x) - h_{l0}(x)) \to N(0, H^{-1} + H^{-1}\Lambda^\star H^{-1})$. $\square$

Proof of Theorem 3.3.7 Theorem 3.3.7 follows immediately with Theorem 3.3.4, Lemma B.0.11 and the Slutsky theorem. $\square$

APPENDIX C

APPENDIX OF CHAPTER 4

Numerical solutions of the price dividend ratio This appendix describes the algorithm that I used to solve the price-dividend ratio. From equation 4.3, the price-dividend ratio ω_t is a function of current consumption growth rate z_t and the subjective expectations as followings:

$$\tilde{\omega}(z_t) = \mathbf{E}^s[\beta(e^{\tilde{z}_{t+1}}(\tilde{\omega}(\tilde{z}_{t+1}) + 1)|\boldsymbol{I_t}] \tag{C.1}$$

where z_t follows a distorted distribution: $z_{t+1} - \mu = \Gamma(z_t - \mu) + \tilde{\epsilon}_{t+1}, \ \tilde{\epsilon} \sim i.i.d \quad N(0, \delta_e^2)$.

I follow the lead of Collard and Juillard (2001) to approximate the price-dividend ratio $\tilde{\omega}_t$ in equation 4.3 numerically. Since the algorithm of the price-dividend ratio under differently lagged models are the same, for simplicity here, I only demonstrate the numerical solutions for AR(1) and ARCH(1) models here. Equation 4.3 can be rewritten as:

$$z_t = h(z_{t-1}, \epsilon_t) \tag{C.2}$$

$$\omega_t = \mathbf{E}_t^s(g(\omega_{t+1}, z_{t+1})) \tag{C.3}$$

Where $\mathbf{E}^s$ is investor's subjective distorted belief. $g(\omega, z) = \beta exp(\theta z)(1 + \omega)$, and $h(z, \epsilon) = (1 - \Gamma)\mu + \Gamma z + \epsilon$, $\epsilon \sim i.i.d \quad N(0, \delta_e^2)$. I'm searching for a function f, such that $\omega_t = f(z_t)$. That is to say,

$$f(z_t) = \mathbf{E}_t^s[g(f(h(z_t, \epsilon_{t+1})), h(z_t, \epsilon_{t+1}))] = \mathbf{E}_t^s[G(z_t, \epsilon_{t+1})] \tag{C.4}$$

I define the error term $H(z_t, \epsilon_{t+1}) = f(z_t) - \mathbf{E}_t^s[G(z_t, \epsilon_{t+1})]$, which becomes the new target of the whole algorithm. We first compute the steady state z^* and ω^*, by solving the following equation:

$$z^* = (1 - \Gamma)\mu + \Gamma z^* \tag{C.5}$$

$$\omega^* = g(\omega^*, z^*) \tag{C.6}$$

Thus, $z^* = \mu$ and $\omega^* = \frac{\beta exp(\theta\mu)}{1-\beta exp(\theta\mu)}$.

Using Taylor expansion of H around the point $(z^*, 0)$ to the second moments, we have:

$$\begin{aligned}
H(z_t, \epsilon_{t+1}) \approx\ & H(z^*, 0) + H_z(z^*, 0)\hat{z}_t + H_\epsilon(x^*, 0)\epsilon_{t+1} \\
& + \frac{1}{2}H_{zz}(z^*, 0)\hat{z}_t^2 + \frac{1}{2}H_{\epsilon\epsilon}(z^*, 0)\epsilon_{t+1}^2 + H_{z\epsilon}(z^*, 0)z_t\epsilon_{t+1} \\
& = [f(z^*) - \mathbf{E}_t^s(G(z^*, 0))] + [f_z(z^*)\hat{z}_t - \mathbf{E}_t^s(G_x(x^*, 0))\hat{x}_t)] \\
& + [f_\epsilon(z^*)\epsilon_{t+1} - \mathbf{E}_t^s(G_\epsilon(G_\epsilon(z^*, 0))\epsilon_{t+1})] \\
& + [\frac{1}{2}(f_{zz}(z^*)\hat{z}_t{}^2 - \mathbf{E}_t^s(G_{zz}(z^*, 0)))\hat{z}_t^2] \\
& + [\frac{1}{2}(-\mathbf{E}_t^s(G_{\epsilon\epsilon}(z^*, 0)))\hat{\epsilon}_{t+1}^2] + [-G_{z\epsilon}(z^*, 0)\hat{z}_t\epsilon_{t+1}] \\
& = f_0 + f_1\hat{z}_t + \frac{1}{2}\hat{z}_t^2 - \mathbf{E}_t^s[G_{0,0} + G_{1,0}\hat{z}_t + G_{0,1}\epsilon_{t+1} + \frac{1}{2}G_{2,0}\hat{z}_t^2 + \frac{1}{2}G_{0,2}\epsilon_{t+1}^2 + G_{1,1}\hat{z}_t\epsilon_{t+1}] \\
& = f_0 + f_1\hat{z}_t + \frac{1}{2}\hat{z}_t^2 - \mathbf{E}_t^s[G_{(0,0)} + G_{1,0}\hat{z}_t + \frac{1}{2}G_{2,0}\hat{z}_t^2 + \frac{1}{2}G_{0,2}\epsilon_{t+1}^2]
\end{aligned} \tag{C.7}$$

where $f_k = \frac{d^k}{dz^k}f(z)|_{z=z^*}$, $\quad G_{i,j} = \frac{\partial^{i+j}}{\partial z^i \partial\epsilon^j}G(z, \epsilon)|_{z=z^*,\epsilon=0}$ and $\hat{z}_t = z_t - z^*$. Equating coefficients allows us to rewrite equation C.7 as:

$$f_0 + f_1\hat{z}_t + \frac{1}{2}f_2\hat{z}_t^2 = G_{(0,0)} + G_{(1,0)}\hat{z}_t + \frac{1}{2}G_{(2,0)}\hat{z}_t^2 + \frac{\delta_e^2}{2}G_{(0,2)} \tag{C.8}$$

Thus by solving f_0, f_1, f_2 defined above via a linear system, we have

$$\tilde{\omega}_t \approx f(z_t) = f_0 + f_0 + f_1 \hat{z}_t + \frac{1}{2} f_2 \hat{z}_t^2 \tag{C.9}$$

where $f_0 = G_{(0,0)} + \frac{\delta_e^2}{2} G_{(0,2)}, \quad f_1 = G_{1,0}$, and $f_2 = G_{2,0}$.

Similarly, for the volatility clustering distorted belief ARCH(1) model, we can approximate the price-dividend ratio by

$$\bar{\omega}_t \approx \bar{f}(z_t) = \bar{f}_0 + \bar{f}_0 + \bar{f}_1 \hat{z}_t + \frac{1}{2} \bar{f}_2 \hat{z}_t^2 \tag{C.10}$$

where $\bar{f}_0 = G_{(0,0)} + \frac{\alpha_0}{2} G_{(0,2)}$, $\bar{f}_1 = G_{1,0}$, and $\bar{f}_2 = G_{2,0} + \alpha_1 G_{(0,2)}$.

Proof of propositions

Proof. Under the extrapolation bias in mean AR(1) process, the cumulative density function F^s of $\tilde{z}_{t+1}$ is $\tilde{z}_{t+1} \sim N(\mu + \Gamma(z_t - \mu), \delta^2)$. The true process of z_{t+1} is i.i.d $N(\mu_r, \delta_r^2)$, and has cumulative density function F. Measure dF^s is absolutely continuous with respect to Lebesgue measure dF, and dF^s is also a finite measure with σ-algebra $\mathbf{M}(\infty, \infty)$, by Radon-Nikodym Theorem, there is a non-negative integrable function h on (∞, ∞), which is unique up to measure 0, such that

$$F^s(A) = \int_A h dF \tag{C.11}$$

We assume z_{t+1} follows AR(P) process, then substituting the function form of subjective measure F^s and true process measure F on both hand sides, we have the following equation:

$$\int_A \frac{1}{\sqrt{2\pi}\delta_d} e^{\frac{-(z_{t+1}-\mu_d-\sum_{j=1}^{P}\Gamma_j(z_{t-j}-\mu_d))^2}{2\delta_d^2}} dz_{t+1} = \int_A h \bullet \frac{1}{\sqrt{2\pi}\delta_r} e^{-\frac{(z_{t+1}-\mu_r)^2}{2\delta_r^2}} dz_{t+1} \tag{C.12}$$

$$h = \frac{\delta_r}{\delta_d} e^{\frac{(z_{t+1}-\mu_r)^2}{2\delta_r^2} - \frac{(z_{t+1}-\mu_d-\sum_{j=1}^{P}\Gamma_j(z_{t-j}-\mu_d)))^2}{2\delta_d^2}} \tag{C.13}$$

The subjective expectation with respect to the subjective measure F^s then can be rewritten as

$$\begin{aligned} \mathbf{E}^s(f(z_{t+1})) &= \int_{-\infty}^{\infty} (g(z_{t+1}))dF^s \\ &= \int_{-\infty}^{\infty} g(z_{t+1}) \bullet h(z_{t+1})dF \\ &= \mathbf{E}(g(z_{t+1}) \bullet h(z_{t+1})) \end{aligned} \tag{C.14}$$

□

BIBLIOGRAPHY

Andrew B Abel. Asset prices under habit formation and catching up with the joneses. Technical report, National Bureau of Economic Research, 1990.

Yacine Aït-Sahalia. Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica*, 70(1):223–262, 2002.

Yacine Ait-Sahalia, Jianqing Fan, and Heng Peng. Nonparametric transition-based tests for jump diffusions. *Journal of the American Statistical Association*, 2012.

Donald WK Andrews. Asymptotic normality of series estimators for nonparametric and semiparametric regression models. *Econometrica: Journal of the Econometric Society*, pages 307–345, 1991.

S Borağan Aruoba, Jesús Fernández-Villaverde, and Juan F Rubio-Ramirez. Comparing solution methods for dynamic equilibrium economies. *Journal of Economic dynamics and Control*, 30(12):2477–2508, 2006.

Patrick Bajari, C Lanier Benkard, and Jonathan Levin. Estimating dynamic models of imperfect competition. *Econometrica*, 75(5):1331–1370, 2007.

James Banks, Richard Blundell, and Agar Brugiavini. Risk pooling, precautionary saving and consumption growth. *The Review of Economic Studies*, 68(4):757–779, 2001.

Ravi Bansal and Amir Yaron. Risks for the long run: A potential resolution of asset pricing puzzles. *The Journal of Finance*, 59(4):1481–1509, 2004.

Ravi Bansal, Dana Kiku, and Amir Yaron. An empirical evaluation of the long-run risks model for asset prices. *Critical Finance Review*, 1 (1):183–221, 2012. ISSN 2164-5744. doi: 10.1561/104.00000005. URL `http://dx.doi.org/10.1561/104.00000005`.

Nicholas Barberis and Ming Huang. Preferences with frames: a new utility specification that allows for the framing of risks. *Journal of Economic Dynamics and Control*, 33(8):1555–1576, 2009.

Nicholas Barberis, Ming Huang, and Tano Santos. Prospect theory and asset prices. Technical report, National bureau of economic research, 1999.

Nicholas Barberis, Robin Greenwood, Lawrence Jin, and Andrei Shleifer. X-capm: An extrapolative capital asset pricing model. *Journal of Financial Economics*, 115 (1):1–24, 2015.

Robert J Barro. Rare disasters and asset markets in the twentieth century. *The Quarterly Journal of Economics*, 121(3):823–866, 2006.

Robert J Barro. Rare disasters, asset prices, and welfare costs. *The American Economic Review*, 99(1):243–264, 2009.

Jason Beeler and John Y Campbell. The long-run risks model and aggregate asset prices: an empirical assessment. Technical report, National Bureau of Economic Research, 2009.

Olivier Blanchard. Do dsge models have a future? Technical report, 2016.

Bruce M Brown. Martingale central limit theorems. *The Annals of Mathematical Statistics*, 42(1):59–66, 1971.

Martin Browning, Lars Peter Hansen, and James J Heckman. Micro data and general equilibrium models. *Handbook of macroeconomics*, 1:543–633, 1999.

Craig Burnside. Solving asset pricing models with gaussian shocks. *Journal of Economic Dynamics and Control*, 22(3):329–340, 1998.

Ricardo J Caballero. Consumption puzzles and precautionary savings. *Journal of monetary economics*, 25(1):113–136, 1990.

Ovidiu L Calin, Yu Chen, Thomas F Cosimano, and Alex A Himonas. Solving asset pricing models when the price–dividend function is analytic. *Econometrica*, 73(3): 961–982, 2005.

John Y Campbell. Consumption-based asset pricing. *Handbook of the Economics of Finance*, 1:803–887, 2003.

John Y Campbell and John H Cochrane. By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of political Economy*, 107 (2):205–251, 1999.

John Y Campbell and John H Cochrane. *Explaining the poor performance of consumption-based asset pricing models*, volume 55. Wiley Online Library, 2000.

John Y Campbell and Robert J Shiller. The dividend-price ratio and expectations of future dividends and discount factors. *Review of financial studies*, 1(3):195–228, 1988.

Stephen Cecchetti and Pok Lam. Asset pricing with distorted beliefs: Are equity returns too good to be true? *American Economic Review*, 90(4):787–805.

Stephen G Cecchetti, Pok-Sang Lam, and Nelson C Mark. Asset pricing with distorted beliefs: Are equity returns too good to be true? *The American Economic Review*, 2000.

Bin Chen and Yongmiao Hong. Testing for smooth structural changes in time series models via nonparametric regression. *Econometrica*, 80(3):1157–1183, 2012.

Xiaohong Chen, Jack Favilukis, and Sydney C Ludvigson. An estimation of economic models with recursive preferences. *Quantitative Economics*, 4(1):39–83, 2013.

Yu Chen, Thomas F Cosimano, and Alex A Himonas. Analytic solving of asset pricing models: The by force of habit case. *Journal of Economic Dynamics and Control*, 32(11):3631–3660, 2008.

Yin-Wong Cheung and Lilian K Ng. International evidence on the stock market and aggregate economic activity. *Journal of empirical finance*, 5(3):281–296, 1998.

Lawrence J Christiano. Solving the stochastic growth model by linear-quadratic approximation and by value-function iteration. *Journal of Business & Economic Statistics*, 8(1):23–26, 1990.

Michael T Cliff. Asset pricing tests under price-dividend ratio restrictions. 2001.

John H Cochrane. Explaining the variance of price–dividend ratios. *Review of Financial Studies*, 5(2):243–280, 1992.

John H Cochrane. *Asset Pricing:(Revised Edition)*. Princeton university press, 2009.

Andrew J Culham. Computational techniques for solving asset pricing models. Technical report, Working Papers 1, Florida State University, Department of Mathematics, 2005.

Robert M De Jong. A note on "convergence rates and asymptotic normality for series estimators": uniform convergence rates. *Journal of Econometrics*, 111(1): 1–9, 2002.

Wouter J Den Haan and Albert Marcet. Accuracy in simulations. *The Review of Economic Studies*, 61(1):3–17, 1994.

Stephen G Donald and Whitney K Newey. A jackknife interpretation of the continuous updating estimator. *Economics Letters*, 67(3):239–243, 2000.

Hilmar Drygas. Gauss-markov estimation for multivariate linear models with missing observations. *The Annals of Statistics*, pages 779–787, 1976a.

Hilmar Drygas. Weak and strong consistency of the least squares estimators in regression models. *Probability Theory and Related Fields*, 34(2):119–127, 1976b.

Martin S Eichenbaum, Lars Peter Hansen, and Kenneth J Singleton. A time series analysis of representative agent models of consumption and leisure choice under uncertainty. *The Quarterly Journal of Economics*, 103(1):51–78, 1988.

Larry G Epstein and Stanley E Zin. Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework. *Econometrica: Journal of the Econometric Society*, pages 937–969, 1989.

Eugene F Fama. Stock returns, expected returns, and real activity. *The Journal of Finance*, 45(4):1089–1108, 1990.

Jianqing Fan and Irene Gijbels. *Local polynomial modelling and its applications: monographs on statistics and applied probability 66*, volume 66. CRC Press, 1996.

Jesús Fernández-Villaverde and Juan F Rubio-Ramírez. Solving dsge models with perturbation methods and a change of variables. *Journal of Economic Dynamics and Control*, 30(12):2509–2531, 2006.

Jesús Fernández-Villaverde and Juan F Rubio-Ramírez. Estimating macroeconomic models: A likelihood approach. *The Review of Economic Studies*, 74(4):1059–1087, 2007.

Jesús Fernández-Villaverde, Juan F Rubio Ramírez, and Frank Schorfheide. Solution and estimation methods for dsge models. Technical report, National Bureau of Economic Research, 2016.

Patricia Fraser and Oluwatobi Oyefeso. Us, uk and european stock market integration. *Journal of Business Finance & Accounting*, 32(1-2):161–181, 2005.

A Ronald Gallant. On the bias in flexible functional forms and an essentially unbiased form: the fourier flexible form. *Journal of Econometrics*, 15(2):211–245, 1981.

A Ronald Gallant and Geraldo Souza. On the asymptotic normality of fourier flexible form estimates. *Journal of Econometrics*, 50(3):329–353, 1991.

A Ronald Gallant and George Tauchen. Seminonparametric estimation of conditionally constrained heterogeneous processes: Asset pricing applications. *Econometrica: Journal of the Econometric Society*, pages 1091–1120, 1989.

A Ronald Gallant and George Tauchen. Which moments to match? *Econometric Theory*, pages 657–681, 1996.

Jiti Gao and Han Hong. A computational implementation of gmm. *Available at SSRN 2503199*, 2015.

Robin Greenwood and Andrei Shleifer. Expectations of returns and expected returns. *Review of Financial Studies*, 27(3):714–746, 2014.

Chirok Han and Peter C. B. Phillips. Gmm with many moment conditions. *Econometrica*, 74(1):147–192, 2006. ISSN 00129682, 14680262. URL `http://www.jstor.org/stable/3598926`.

Q Han and YM Hong. National industrial policy asset price and investor behavior. *Economics Research*, 12:143–158, 2014.

Lars Peter Hansen and José A Scheinkman. Recursive utility in a markov environment with stochastic growth. *Proceedings of the National Academy of Sciences*, 109(30):11967–11972, 2012.

Lars Peter Hansen and Kenneth J Singleton. Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica: Journal of the Econometric Society*, pages 1269–1286, 1982.

Lars Peter Hansen, John Heaton, and Amir Yaron. Finite-sample properties of some alternative gmm estimators. *Journal of Business & Economic Statistics*, 14(3): 262–280, 1996.

Yongmiao Hong and Halbert White. Consistent specification testing via nonparametric series regression. *Econometrica: Journal of the Econometric Society*, pages 1133–1159, 1995.

Yongmiao Hong, Haitao Li, and Feng Zhao. Out-of-sample performance of discrete-time spot interest rate models. *Journal of Business & Economic Statistics*, 2012.

V Joseph Hotz and Robert A Miller. Conditional choice probabilities and the estimation of dynamic models. *The Review of Economic Studies*, 60(3):497–529, 1993.

Jianhua Z Huang. Local asymptotics for polynomial spline regression. *The Annals of Statistics*, 31(5):1600–1635, 2003.

Ravi Jagannathan and Srikant Marakani. Price-dividend ratio factor proxies for long-run risks. *Review of Asset Pricing Studies*, 5(1):1–47, 2015.

Kenneth L Judd. Projection methods for solving aggregate growth models. *Journal of Economic Theory*, 58(2):410–452, 1992.

Kenneth L Judd. *Numerical methods in economics*. MIT press, 1998.

Michel Juillard. Local approximation of dsge models around the risky steady state. Technical report, Mimeo October, 2011.

Jinill Kim and Sunghyun Henry Kim. Spurious welfare reversals in international business cycle models. *journal of International Economics*, 60(2):471–500, 2003.

Hongquan Li and Yongmiao Hong. Financial volatility forecasting with range-based autoregressive volatility model. *Finance Research Letters*, 8(2):69–76, 2011.

Zheng Liu, Pengfei Wang, and Tao Zha. Land-price dynamics and macroeconomic fluctuations. *Econometrica*, 81(3):1147–1184, 2013.

Robert E Lucas Jr. Asset prices in an exchange economy. *Econometrica: Journal of the Econometric Society*, pages 1429–1445, 1978.

Daniel McFadden. A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica: Journal of the Econometric Society*, pages 995–1026, 1989.

Costas Meghir and Luigi Pistaferri. Income variance dynamics and heterogeneity. *Econometrica*, 72(1):1–32, 2004.

Rajnish Mehra and Edward C Prescott. The equity premium: A puzzle. *Journal of monetary Economics*, 15(2):145–161, 1985.

Whitney K Newey. Uniform convergence in probability and stochastic equicontinuity. *Econometrica: Journal of the Econometric Society*, pages 1161–1167, 1991.

Whitney K Newey. Series estimation of regression functionals. *Econometric Theory*, 10(01):1–28, 1994.

Whitney K Newey. Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics*, 79(1):147–168, 1997.

Whitney K Newey. Many weak moment asymptotics for the continuously updated gmm estimator. In *MIT Working Paper*. 2004.

Whitney K Newey and James L Powell. Instrumental variable estimation of nonparametric models. *Econometrica*, 71(5):1565–1578, 2003.

Whitney K Newey and Richard J Smith. Higher order properties of gmm and generalized empirical likelihood estimators. *Econometrica*, 72(1):219–255, 2004.

Walter Pohl, Karl Schmedders, and Ole Wilms. Solving asset-pricing models with recursive preferences. 2014.

David Pollard. *Convergence of stochastic processes*. Springer Science & Business Media, 2012.

Stephen Portnoy. Asymptotic behavior of m estimators of p regression parameters when p2/n is large; ii. normal approximation. *The Annals of Statistics*, pages 1403–1417, 1985.

Thomas A Rietz. The equity risk premium a solution. *Journal of monetary Economics*, 22(1):117–131, 1988.

Manuel S Santos. Accuracy of numerical solutions using the euler equation residuals. *Econometrica*, 68(6):1377–1402, 2000.

Stephanie Schmitt-Grohe and Martın Uribe. Solving dynamic general equilibrium models using a second-order approximation to the policy function. *Journal of economic dynamics and control*, 28(4):755–775, 2004.

Robert J Shiller, Fumiko Kon-Ya, and Yoshiro Tsutsui. Investor behavior in the october 1987 stock market crash: The case of japan. *Journal of the Japanese and International Economies*, 5(1):1–13, 1991.

James H Stock and Jonathan H Wright. Asymptotics for gmm estimators with weak instruments, 1996.

Thomas D Tallarini. Risk-sensitive real business cycles. *Journal of Monetary Economics*, 45(3):507–532, 2000.

George Tauchen and Robert Hussey. Quadrature-based methods for obtaining approximate solutions to nonlinear asset pricing models. *Econometrica: Journal of the Econometric Society*, pages 371–396, 1991.

John B Taylor and Harald Uhlig. Solving nonlinear stochastic growth models: A comparison of alternative solution methods. *Journal of Business & Economic Statistics*, 8(1):1–17, 1990.

Mark P Taylor and Ian Tonks. The internationalisation of stock markets and the abolition of uk exchange control. *The Review of Economics and Statistics*, pages 332–336, 1989.

Efthymios G Tsionas. Exact solution of asset pricing models with arbitrary shock distributions. *Journal of Economic Dynamics and Control*, 27(5):843–851, 2003.

Sara A Van de Geer. *Applications of empirical process theory*, volume 91. Cambridge University Press Cambridge, 2000.

Jessica A Wachter. Can time-varying risk of rare disasters explain aggregate stock market volatility? *The Journal of Finance*, 68(3):987–1035, 2013.

Michael Woodford. Inflation stabilization and welfare. *Contributions in Macroeconomics*, 2(1), 2002.

Joachim Zietz. Log-linearizing around the steady state: A guide with examples. *Available at SSRN 951753*, 2006.